

稲垣 博人 大森 信行 田中 一男

NTT サイバースリユーション研究所

## 1 はじめに

携帯電話や携帯端末の低価格化、多機能化により、多くの人が携帯電話や携帯端末を持ち歩くようになり、携帯電話／携帯端末がコンシューマ端末として育ってきた。特に、今まで、デスクトップのパソコンだけのインターネットの世界が、携帯端末の多機能化により、多くの携帯端末でもアクセス可能となってきた。もちろん、インターネットの世界は、デスクトップパソコンだけのものではないが、インターネットの世界を見るためには、PCの設定やISPへのリモートアクセスなどいくつかのハードルがあった。一方で、携帯電話や携帯端末は、プロトコルとしてデスクトップパソコンと同様なIP (Internet Protocol) を標準的に採用し、インターネットの世界をデスクトップと同様に享受することが可能になったものの、携帯端末の表現能力の問題により、インターネットのすべての世界を利用するというよりも、デスクトップはデスクトップの世界、携帯端末は、携帯端末の世界となってしまっている。これは、PCで見るとに適したコンテンツは、必ずしも携帯端末で見るとに適したコンテンツではないということに起因している。そこで、本稿では、このような携帯端末でプレゼンテーションするのに適していないインターネットコンテンツを携帯端末で短時間で見やすくプレゼンテーションするようにコンテンツを変換する携帯端末向けコンテンツ速覧手法について述べる。特に、コンテンツ対象としては、インターネットのHTMLコンテンツを対象とし、インターネットを閲覧できる携帯電話等の携帯端末で、(迅)速に(閱)覧可能なように、コンテンツをブロックに分割し、個々のブロックにおける話題を抽出し、話題をリンクとして表示する速覧手法を検討する。

## 2 携帯端末向けのコンテンツ速覧

インターネット上のコンテンツであるHTML文書は、PCやWSなどプレゼンテーションする能力が高い機器向けに作成されたコンテンツである。そのため、PCなどに比べ、低機能で、プレゼンテーションに利用するリソース(表示面積、レンダリング能力、メモリなど)が極めて少ない携帯端末等では、コンテンツを閲覧することができない。例えば、初期のi-mode携帯電話

では、表示可能色は、白黒、さらに表示可能なコンテンツのサイズは、2Kbyte以下を推奨し、画像はGIFのみ、などの極めて多くの制限がある。このような、リソースに制限がある携帯端末に対してPCなどのようにプレゼンテーション能力の高い機器向けのコンテンツを表示するためにはコンテンツ変換・表示が必須である。HTMLコンテンツを対象とする場合、以下の制限をすべて満たすようにコンテンツを変換・表示しなければならない。

- 蓄積可能なメモリサイズ
- タグ・コンテンツに関する処理
  - － 処理可能なタグ・コンテンツ
  - － 処理すべきタグ・コンテンツ
  - － 削除すべきタグ・コンテンツ
- 表示可能なメディア
  - － 画像 (GIF, JPEG, など)
  - － 音声 (AIFF, WAVE など)
  - － 映像 (MPEG, AVI など)
- 処理可能なスクリプト (JavaScript など)

これらの制限は、相互に関係する場合もある。

コンテンツを変換・表示するためのコンテンツ速覧手法としては、大きく分けて3種類が考えられる。

- (a) コンテンツをブロック化し、ブロックの特徴を表す要旨ごとにリンク表示する手法 (Indicative コンテンツ速覧手法)
- (b) コンテンツの特徴的な情報をリンク表示する手法
- (c) コンテンツを表示可能な部分ごとに表示する手法

我々は、ユーザがコンテンツ中から自分に必要な情報を短時間で選択し、閲覧するタスクを行なうのに優れた(a)の手法を採用した。(b)の手法は、i-GATE<sup>1)</sup>等で利用されている方法である。i-GATEの場合、HTMLコンテンツの特徴的な情報として最も有効であるハイパーリンクを中心にリンク表示する手法である。(b)のような手法では、確かに特徴的な情報(ハイパーリンクなど)を中心にWEBを見歩く場合には適しているものの、HTMLコンテンツの中身を読むのに特に適しているわけではない。一方、(c)の手法では、表示可能な部分ごとに順次表示するため、HTMLコンテンツを順次見ていく場合には適しているものの、速覧性に劣っている。

A study of internet contents skimming method for private mobile terminals.

Hirohito INAGAKI, Nobuyuki OHMORI,  
and Kazuo TANAKA.

NTT Cyber Solutions Laboratories

### 3 Indicative コンテンツ速覧手法

Indicative コンテンツ速覧手法では、入力されたコンテンツを、適切な単位でブロック化し、ブロックの特徴を表す要旨ごとにコンテンツをリンク表示する手法である。この手法は要約<sup>2)</sup>などでも用いられている。本手法では、以下の処理を行なう。

- (1) コンテンツの読み取り処理
- (2) コンテンツのブロック化処理
- (3) ブロック化コンテンツからの要旨抽出処理
- (4) 各ブロックを表す要旨リンク表示処理

コンテンツ読み取り処理では、指定されたコンテンツを読み込む処理である。インターネットであれば、URLが指定され、該当するURLをクローラなどで収集する処理である。

コンテンツのブロック化処理では、収集したHTMLコンテンツのタグを意識し、なるべくタグで囲まれたブロックを分割しないようにすると共に、携帯端末のプレゼンテーションにおける制限(例えば、表示可能な最大容量)に応じて、ブロック化を行なう。

ブロック化コンテンツからの要旨抽出処理では、ブロックの特徴を表す要旨を各ブロックから抽出する処理である。本処理は、ブロック化の処理と同様に、コンテンツ速覧手法の性能を決定付ける重要な処理である。我々は、以下の観点から、ブロックの内容を特徴付けた。

- ブロックの特徴を表す特殊なタグ (H1, FONT, BOLD など)
- ブロックの特徴を表す話題

ブロックの特徴を表す特殊なタグとは、HTMLで記述されたコンテンツは、必ずしもタグがセマンティクスを持つわけではなく、タグがプレゼンテーションのための手段として利用されていることが多い。FONTの大きさや属性を変更するのは、通常のセマンティクスでは、大きいものほど、FONTの属性が標準でないものほど要旨に関連する特徴的な情報が表示される。しかし、通常のセマンティクス以外で利用されている場合もあるので、セマンティクスが適切であるか、特徴的であるかをHTMLコンテンツ全体からみて判断し、ブロックの特徴を表す特殊なタグである場合、該タグに囲まれている文字列を要旨とした。

ブロックの特徴を表す話題を抽出する手法としては、我々は、“快速覧システム<sup>3)</sup>”を利用した。快速覧とは、対象テキストに対して形態素解析を行ない、形態素解析により抽出された、キーワード(たとえば、“まず”、“第一に”、“例えば”)などから、ある文章中に記述されている話題(主題)を抽出するシステムであ

る。実時間程度の処理速度を実現するために、係り受け解析や意味解析などは行わず、形態素解析のみを実行する。形態素解析としては、InfoBee 形態素解析エンジン<sup>4)</sup>を用い、InfoBee 形態素解析エンジンに組み込んだ形で使った。もちろん、特殊タグや快速覧を使ってもブロックの中から適切な要旨を抽出できない場合は、対象とするブロックから適切な単語列を抽出する。

各ブロックを表す要旨リンク表示処理では、(3)の処理で抽出された各ブロックの要旨を表示し、要旨を表す文字列に対してハイパーリンクを設定し、対象とするブロック全体を表示する処理である。最終的に、ユーザは、表示された要旨を見て、クリックすることにより、必要とするHTMLコンテンツのブロックを閲覧できる仕組みである。

### 4 検索エンジンとの連動

Indicative コンテンツ速覧手法を用いることにより、インターネットの検索を携帯端末で行なうことも可能となる。例えば、検索エンジンとしてgoo<sup>5)</sup>などを利用することを考えると、検索処理結果に対して、本コンテンツ速覧手法を介してコンテンツを表示するようにすれば、どんなコンテンツでも閲覧可能となる。

### 5 まとめ

本稿では、インターネットのHTMLコンテンツをタグ情報をもとに分割し、分割したブロックの中からブロックの内容を表すに適した話題を表示することにより、インターネット上の約2000万URLと言われている膨大なコンテンツを携帯端末でも検索・閲覧することができるコンテンツ速覧手法について述べた。現在、本コンテンツ速覧手法をUNIX WS上に構築した。今後は本手法の有効性を評価する予定である。

#### 参考文献

- 1) <http://www.i-greet.com>.
- 2) H. Saggion and G. Lapalme. The generation of abstracts by selective analysis. *AAAI Press. Technical Report SS-98-06*, 1998.
- 3) 稲垣博人, 早川和宏, 田中一男. 話題構造および文意味構造に基づく文書可視提示方式の提案. 情報処理学会第57回全国大会, 4R-10, 1998.
- 4) 井上孝史, 大久保雅且, 杉崎正之. InfoBee テキスト情報検索技術. NTT R&D ジャーナル, Vol.46, No.10, pp. 1103-1108, 1997.
- 5) <http://www.goo.ne.jp>.