

Web コンテンツ読み上げシステム - Web ダイナミック コンテンツ音声インタフェース -

2J-4

大森 信行 榎原 佳織 宮本 勝 稲垣 博人
日本電信電話株式会社 NTT サイバーソリューション研究所

1 はじめに

インターネットにおいて、検索、コミュニティなど Web を利用した種々のサービスがポータルサイトで運営されている。これらのサービスは、PC の利用を前提としており、キーボードあるいはマウスなどポインティングデバイスにより情報を入力するため、ユーザは PC の利用方法を習得している必要がある。また、i モードのように電子メールや Web 閲覧などのインターネット接続機能を持った携帯電話端末は、主要な入力手段がテンキーのみに制限されている。そこで、本稿では非 PC ユーザも利用可能なポータルサイトを目的として、音声インタフェースを持つボイスポータル・サービスを提案する。従来は、固定的な情報を読み上げるタイプが多かったが、我々の提案するシステムでは、ユーザが Web コンテンツの中から、利用するコンテンツを自由に選択し、ユーザの指定した部分のテキストを音声によって読み上げるサービスを提案する。

2 WWW コンテンツ読み上げシステム

本稿で目指す音声インタフェースを持つコンテンツ読み上げサービスは、ユーザが指定したコンテンツの特定部分を音声合成により読み上げ、音声メッセージをユーザに対して出力するサービスである。

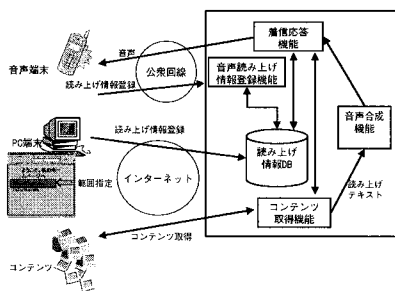


図 1: WEB コンテンツ読み上げシステム

A study on speech interface for WWW contents.
Nobuyuki OHMORI, Kaori NARAHARA,
Masaru MIYAMOTO, and Hirohito INAGAKI
NTT Cyber Solutions Laboratories, NTT Corporation

本システムの概要を図 1 に示す。本システムは、以下の機能から構成されている。

- (1) 音声合成機能
- (2) コンテンツ取得機能
- (3) 読み上げ情報 DB
- (4) 着信応答機能
- (5) 読み上げ情報登録機能
- (6) 音声読み上げ情報登録機能

本システムの利用者は、あらかじめ読み上げ情報登録として、読み上げを希望するコンテンツの特定部分を読み上げ情報 DB に登録する。

2.1 音声合成機能

ユーザが指定したコンテンツの特定部分のテキストを音声により読み上げるための音声出力インタフェースである。入力にはコンテンツ取得機能が出力したテキストであり、出力はそのテキストを読み上げた音声信号である。合成された音声信号は、公衆回線等を経由してユーザの端末から出力される。

2.2 コンテンツ取得機能

コンテンツ取得機能は、ユーザの指定した WWW コンテンツを取得し、そのコンテンツ内でユーザが読み上げを希望する部分、つまり読み上げ部分を特定し、その部分を読み上げテキストとして出力する。

WWW のコンテンツで一般的にユーザが頻繁に見る情報、例えばニュース、天気予報、株価などのコンテンツは更新頻度が他と比べて非常に大きく、数時間単位で更新されることも多い。このような特徴から、読み上げを登録したコンテンツがユーザが読み上げさせようとアクセスした段階ですでに更新されている可能性が大きい。つまり、読み上げ部分を正確に特定するためには、読み上げ情報の登録時と内容が変更されていてもユーザが指定した個所を適切に特定する必要がある。

読み上げ部分の特定に着目すると、コンテンツの更新は大きく 2 つに分類することができる。

- (1) 読み上げ部分の周囲が変更されている場合
- (2) 読み上げ部分に変更されている場合

前者については、ページ中のバナー広告などがアクセスする度に变化するような場合である。読み上げ部分を特定する際に、コンテンツの先頭からのバイト数によって指定することは難しい。そのため、読み上げ情報 DB 中の初期登録文字列と一致する部分のみを読み上げテキストとすることで、読み上げ部分を特定することが可能である。

後者の例としては、ユーザが天気予報のページ中の横浜地方の天気を見ようとして「横浜 明日 晴れ」という部分を初期登録文字列として登録している場合、「晴れ」が「曇り」や「雨」に変わる可能性がある。このような場合は、初期登録文字列との一致度がもっとも大きな部分を読み上げ部分として選択する。

一度度の計算方法としては、以下の通りである。

- (1) コンテンツから視覚情報の記述に利用されるタグを取り除く。
- (2) HTML タグ情報により、TABLE 内の位置などのコンテンツ内の大まかな位置を特定する。
- (3) テキストから、初期登録文字列と最長一致する部分を取り出す。このとき取り出す読み上げ部分は、最長一致する部分の先頭から、初期登録文字列の長さまでの範囲である。

読み上げ部分の特定に HTML タグを手がかりとして利用できる。HTML タグは文字の大きさや色などの視覚表現の記述にも利用される。読み上げ部分特定の精度低下を防ぐため、これらの HTML タグを取り除いて処理を行った。

2.3 読み上げ情報 DB

読み上げ情報 DB には、読み上げ情報に関する情報が登録されている。ユーザが読み上げ情報を一件登録する度に、以下の各項目を一レコードとして登録される。

- (1) レコード ID: DB 内でレコードを特定するための ID。
- (2) 発信者 ID: ユーザの電話番号。ユーザの識別に利用する。
- (3) URL: ユーザの指定したコンテンツのインターネット上の位置を示す URL。
- (4) 初期登録文字列: 読み上げ情報登録時にユーザが指定したコンテンツ内の読み上げ文字列。

発信者 ID、URL、初期登録文字列は、読み上げ情報登録機能により DB に送信される。

2.4 着信応答機能

着信応答機能は、公衆回線を経由したユーザからの着信を受取り、そのユーザが登録した読み上げ情報を、読み上げ情報 DB から取得する。コンテンツの URL と後

述する初期登録文字列をコンテンツ取得部に渡し、コンテンツ取得部の出力を読み上げテキストとして、音声合成機能に渡す。音声合成機能の出力した音声を公衆回線経由でユーザに対して出力する。

2.5 読み上げ情報登録機能

読み上げ情報登録の機能は、PC 端末の機能であり、ユーザが読み上げを希望するコンテンツおよびコンテンツ内の特定個所を指定し、それを読み上げ情報 DB に登録する機能である。ユーザがコンテンツの閲覧中に選択した文字列を「初期登録文字列」として、コンテンツの URL を「URL」として、PC 端末にあらかじめ登録されている電話番号を「発信者 ID」として、読み上げ情報 DB に送信する。

2.6 音声読み上げ情報登録機能

読み上げ情報登録機能は、PC などの非音声端末の操作を必要とする。音声読み上げ情報登録機能は、携帯電話などの音声端末を利用し、読み上げ情報の登録をすべて音声で可能とする。具体的には、以下の処理を行う。

- (1) ユーザからの電話に着信し発信者 ID を取得する。
- (2) ユーザがキーワードを発話し、それを音声認識する。
- (3) 読み上げ情報 DB から、認識したキーワードを初期登録文字列に含むレコードを取り出す。
- (4) 取り出したレコードの初期登録文字列を音声合成により読み上げる。
- (5) ユーザが登録を希望したレコードの初期登録文字列、URL を DB に新規レコードとして登録する。
この際、発信者 ID は (1) で取得した番号とする。

ここで、読み上げ情報 DB の各レコードの初期登録文字列に含まれる単語を音声認識するため、各行の初期登録文字列を形態素解析 [1] し、そこに含まれる単語を取り出し、読み仮名とともに音声認識辞書に登録しておく。

3 まとめ

ユーザが指定した任意の WWW コンテンツを音声で読み上げる音声閲覧サービスを提案した。今後は、ユーザの読み上げ部分の特定手法について、評価を行うとともに、高精度化を検討する予定である。

参考文献

- [1] 井上 他: InfoBee テキスト情報検索技術, NTT R&D No.10, pp.1103-1108, 1997