

携帯環境のための Web 検索結果の動的再構成と受動的視聴

灘 本 明 代[†] 近 藤 宏 行^{††}, 田 中 克 己^{†††}

本研究では、携帯電話に代表されるように、インタラクションや表示領域が大幅に制限されている携帯端末上において、Web の検索結果を最小限のインタラクションにより受動的に視聴できるシステムの提案を行う。具体的には、Web の検索結果の各々のページからカルーセルコンポーネントと呼ぶ画像と音声で表現されるコンポーネントを作成し、ユーザのインタラクション時に、そのカルーセルコンポーネントを類似カルーセル、相違カルーセル、詳細カルーセル、簡潔カルーセルに動的に再構成し、カルーセル呈示する。我々はこのカルーセル呈示システムを WebCarousel と呼ぶ。WebCarousel の特徴は、(1) Web ページの要約であるカルーセルコンポーネントにより構成すること、(2) カルーセルコンポーネントは画像と音声により構成すること、(3) ユーザのインタラクションによる類似カルーセル、相違カルーセル、詳細カルーセル、簡潔カルーセルの動的再構成、(4) カルーセルコンポーネントのカルーセル呈示である。

Dynamic Restructuring and Passive Viewing of Web Search Results for Mobile Environments

AKIYO NADAMOTO,[†] HIROYUKI KONDO^{††}, and KATSUMI TANAKA^{†††}

In the present paper, we will propose a new way of organizing Web search results and of viewing those results passively in the mobile environment which has limited display and limited interaction. Specifically, the system makes carousel componets, that are composed of image and voice, from Web search result. Each time of user's interaction, the system automatically computes sets of more-detailed, more-abstracted, similar, and different pages, respectively and reorganizes them as carrousel. We call this system as WebCarousel. The major features of WebCarousel are (1) WebCarousel are composed of carousel components (2) carousel componets are composed of image and voice and made from Web search result pages (3) Web search result is automatically classified into several groups from the viewpoint of the similarity and details, (4) WebCarousel displays carousel components by carousel. In this way users are easy way to get Web search results data.

1. はじめに

近年、携帯電話の i モードサービスに代表されるように、携帯端末上でインターネットの様々な情報サービスの提供が始まっている。しかしながら携帯端末上では下記の制限があるため、Web コンテンツをテキストベースのコンテンツに変換して送信しているのが

現状である。

- 通信速度の制限
- 表示領域の制限
- インタラクション機能の制限

ここでいうインタラクション機能の制限とは、フルキーボードやマウスがないため、PC と同様の操作ができないことである。

今後、次世代携帯電話方式である IMT2000 など、携帯端末環境における通信速度の向上は十分期待される。それとともに、携帯端末環境では、静止画像のみならず動画画像を音声とともに送受信することが可能となる。画像が容易に送受信できるようになると、Web コンテンツを携帯端末上で閲覧したいという要求が高まることが予想される。それにともない、携帯端末上での Web の検索が必要になると考えられる。

しかしながら、携帯端末では携帯する関係上、表示

[†] 神戸大学大学院自然科学研究科情報メディア科学専攻
Division of Information and Media Sciences, Graduate
School of Science and Technology, Kobe University

^{††} 神戸大学大学院自然科学研究科情報知能工学専攻
Division of Computer and Systems Engineering, Grad-
uate School of Science and Technology, Kobe University

^{†††} 京都大学大学院情報学研究所社会情報学専攻
Division of Social Informatics, Graduate School of In-
formatics, Kyoto University

現在、NHK

Presently with Japan Broadcasting Corporation

領域およびインタラクション機能の向上は、PC と同程度の仕様を期待することはできない。

現在、PC 上での Web の検索結果の表示方法は、検索サイト独自のランキングによる検索結果を、ページのタイトルや URL、内容の一部などを文字情報として列挙しており、PC 上でも、一度に 10 から 20 件の検索結果しか表示できない。ましてや、表示領域の小さい携帯端末ではこれらの検索結果をそのまま表示することは困難である。

また、ユーザは表示された検索結果を読み、ページをめくるようにスクロールやクリックという能動的な操作を行い、欲しい検索結果を取得する方法が一般的である。これに対し、携帯端末ではインタラクション機能が限られているため PC と同様に「読む」「クリックする」「スクロールする」などの能動的な操作をユーザに要求することはできない。そのため、Web の検索結果を「見る」「聞く」といった受動的な方法で取得することが望ましいと考えた。

「見る」「聞く」といった受動的視聴を用いることにより、容易にまた片手間に Web の検索結果を取得することが可能となる。よってこのことは、携帯端末上の Web の検索結果の取得方法として望ましい。

しかしながら、Web の検索結果の取得は、ユーザが欲しい検索結果を指定しなければならないため、ある程度のインタラクションが必要である。

そこで本論文では、携帯端末における Web の検索結果の受動的視聴システムである WebCarousel を提案する。WebCarousel によれば、ユーザは、簡単なインタラクションによって、Web の検索結果を絞り込みながらカーセル表示させることができ、検索結果の受動的な取得が可能になる。一般にカーセルとは、回転木馬（メリーゴーランド）や空港での荷物受取り回転テーブルのことを示す。本論文では同じコンテンツを回転木馬のように繰り返し表示することをカーセル表示という。そのカーセル表示される 1 つ 1 つの検索結果をカーセルコンポーネントと呼ぶ。カーセルコンポーネントは画像と音声で構成されている。また、一度にカーセル表示されるカーセルコンポーネントの集合をカーセルと呼ぶ。そして、ユーザが指示したカーセルコンポーネントと内容が類似、相違、詳細、簡潔するカーセルコンポーネントから再構成されるカーセルを類似カーセル、相違カーセル、詳細カーセル、簡潔カーセルと呼び、そのユーザが行うインタラクション名を各々類似カーセル生成、相違カーセル生成、詳細カーセル生成、簡潔カーセル生成と呼ぶ。

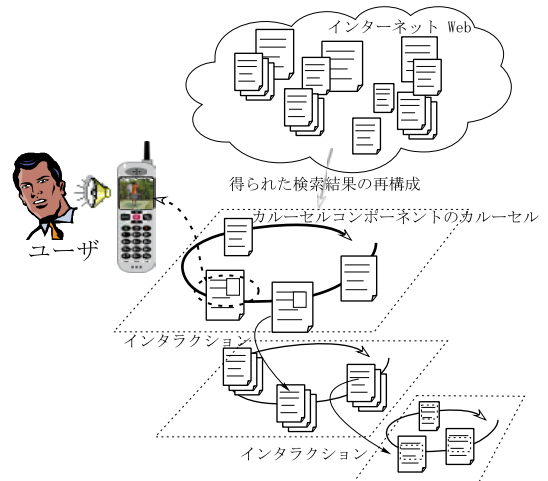


図 1 WebCarousel のイメージ図

Fig. 1 Image of WebCarousel.

具体的に WebCarousel は、以下の手順で検索結果を携帯端末上に表示する。

- 検索結果の各々のページからカーセルコンポーネントと呼ぶ画像と音声で表現されるコンポーネントを作成する。
- 検索結果上位 10 件のカーセルコンポーネントをカーセル表示する。これを初期カーセルと呼ぶ。
- 類似カーセル生成、相違カーセル生成、詳細カーセル生成、簡潔カーセル生成の 4 つのインタラクションを用意し、ユーザのインタラクションにより、検索結果を意味の関係に基づいた類似カーセル、相違カーセル、詳細カーセル、簡潔カーセルに動的に再構成する。
- 再度カーセルコンポーネントを作成し、携帯端末上に表示する。

図 1 に WebCarousel のイメージ図を示す。

一般に Web 情報を携帯端末上で表示するためのシステムとして現在、2 つの方法が用いられている。1 つは携帯小型端末上でインターネットを実現するための通信プロトコルの総称である WAP¹⁾ (Wireless Application Protocol) に代表されるようにプロキシサーバを置く方法であり、もう 1 つは i-モード (compactHTML) に代表されるように携帯端末用にコンテンツを新たに作成する方法である。

プロキシサーバを用いる方法は Web コンテンツを携帯端末用に特別に記述しなくてよいため、Web 作成者の負担にならない。我々はこのプロキシサーバを用いる方法が望ましいと考え、WebCarousel ではプロキシサーバ上で変換する方法を用いた。

以上より、本論文では、

- 携帯端末における検索結果の受動的視聴の方法、
 - カラーセルコンポーネントの生成
 - カラーセルの生成
- 意味的關係に基づいたカラーセルの動的再構成、を提案する。

以降、2章では関連研究について、3章でカラーセルコンポーネントについて、4章でカラーセルの動的再構成を、5章でプロトタイプシステムについて述べ、6章でまとめを述べる。

2. 関連研究

2.1 Web Adaptation

携帯端末環境の急速なる発展にともない、種々な携帯端末に適合させて Web コンテンツを呈示する「Web Adaptation」の技術が注目されている。

平川ら²⁾は、携帯電話に限らず、その他のモバイル端末たとえばカーナビや PDA において統一したコンテンツの提供を行うための、独自のプロキシの実装を行っている。このプロキシは一般の Web コンテンツを各端末のブラウザの機器能力に応じて変換を行い、これにより表示画面の大きさや色数の制限に柔軟に対応できるように作成されている。

スタンフォード大学の Power Browser Project³⁾は既存の Web コンテンツをプロキシサーバを用いて PDA やその他の携帯端末に自動で変換する新しいブラウザマネージャを提案している。このブラウザは、ペン入力による検索やナビゲーションをサポートするために、キーワードによる呈示やアンカー文字列の呈示を行っている。

Horiら⁴⁾は PC や携帯端末上で注釈を用いた Web コンテンツの変換を提唱している。彼らのアプローチは、PC 上と携帯端末上で表示する Web ページの内容を一致させるための additional annotation ファイルと呼ばれるファイルを用い、プロキシサーバを利用して Web ページを PC や携帯端末上で表示する方法である。元の Web ページと彼らの提唱する additional annotation ファイルとのリンクには、XML の XPOINTER/XLINK を用いている。

Maly⁵⁾らはあらかじめ PC 上で指定したパーソナライズした Web 情報を WAP を実装している携帯端末上でテキストベースで呈示する PWPS (Personalized Wireless Portal System) の研究を行っている。

これらの研究は 1 つの Web ページをプロキシサーバを用いて携帯端末上に呈示しているのに対し、本論文では、プロキシサーバを用いるが、携帯端末上に呈

示する対象を検索結果である複数の Web ページとし、その検索結果をカラーセル呈示する点が異なる。

2.2 適合フィードバック

文書検索精度を向上させる手法として、Rocchio の提唱した適合フィードバック⁶⁾がある。Rocchio の適合フィードバックは多くの研究者からその有効性が報告されている。この Rocchio の適合フィードバックは、検索者が入力したキーワードによりシステムが 1 次検索を行い、その結果から妥当と思われるものを検索者自身が選択指示する。そして、その内容をフィードバックすることにより再度検索を行う。実際には、ベクトル空間モデルを用いて、類似文書の検索を行う。本論文では、ベクトル空間モデルを用いて、文書間の類似度の処理を行う点は適合フィードバックと似ている。しかし、適合フィードバックはベクトル空間モデルを用いて、2 次検索を行うのに対し、本研究はベクトル空間モデルを用いてすでに得られた検索結果から呈示するカラーセルを再構成する処理を行うため、基本的に異なる。

2.3 Web の受動的視聴

我々はこれまで、膨大な Web 情報を容易に取得することを目的とし Watch and Listen 型インタフェースである、受動的視聴型番組呈示方式を提唱してきた^{7)~11)}。これらは、テキストや画像で表示されていたコンテンツを、音声やキャラクターアニメーションを用いて表現することで、これまでの Web ブラウザと異なり、インタラクションを極力抑えたうえで有効に情報を提供できる環境を提供している。そこでは、レイアウト情報として与えられている、テキストと画像の時間軸的同期情報を、HTML 文書の木構造を用いて解析し、テキストとそれに対応する画像を含む領域 (同期化可能領域と呼ぶ) を検出する。同期化可能領域の発見は携帯端末においても利用することは有効であると考え、本論文ではこの同期化可能領域の発見手法を基本的に用いて、検索キーワードを含むテキストと同期する画像の発見を行った。

3. カラーセルコンポーネント

3.1 概要

表示領域が小さい携帯端末において、従来の PC のブラウザのように、検索結果を文字情報として列挙し、その検索結果が必要かどうかを判断することは容易ではない。そこで、検索結果として得られたページの中からある一部分を抽出し、それらを画像と音声を用いてカラーセル呈示することを考える。カラーセルの構成要素としてカラーセルコンポーネントを提案する。

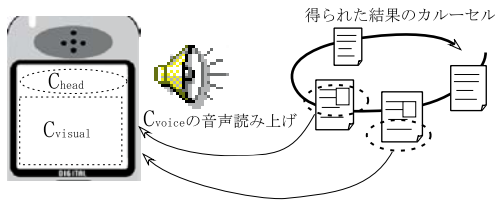


図2 カラーセルコンポーネント
Fig. 2 Carousel Component.

カラーセルコンポーネント CC は検索結果の各ページ P に対応し、ページ P のタイトルと検索キーワードが含まれる文章の一部とその文章と同期する1つの画像からなる。カラーセルコンポーネントは

$$CC = (C_{head}, C_{voice}, C_{visual})$$

のように3つの部分で構成する。

- C_{head} : ヘッダ部分
 P のタイトル, 見出しであり, 画面上にタイトルとして表示される部分を示す。
- C_{voice} : 音声部分
検索キーワードが含まれる P の文章の一部であり, 音声読み上げされる部分を示す。
- C_{visual} : ビジュアル部分
 C_{voice} と同期している画像や文字列である。
 C_{voice} が音声読み上げされている間, 画面上に C_{head} とともに表示される部分を示す。

図2にカラーセルコンポーネントを示す。

カラーセルコンポーネントの生成方法は,

- 初期カラーセルのカラーセルコンポーネント
- 初期カラーセル以降に動的に再構成するカラーセルのカラーセルコンポーネント

の2つがある。このように, C_{voice} の生成方法が異なるため, C_{voice} については, 本章では初期カラーセルのカラーセルコンポーネントの生成方法を述べ, 4.2節で初期カラーセル以降のカラーセルのカラーセルコンポーネントの生成方法を述べる。

カラーセルコンポーネントを構成する各構成部分は Web ページの HTML タグ構造を解析することにより得られる。HTML 文書のタグはその性質上, 以下のような4種類に分けられる。以下では, 解析の際に各タグをどのように利用するかについても述べる。

- 構造タグ ($H1 \sim Hn, P, BLOCKQUOTE, DIV, VL, OL, DL, TABLE$)
これらのタグは主に文書構造を示すために用いられる。一般に HTML から文書構造を抜き出すためにこれらのタグを利用する。
- 強調タグ ($STRONG, EM, TT, I, U, B, BIG, SMALL, STRIKE, S, FONT$)

これらのタグは主に強調などの特殊な表示を行うときなどに用いられる。これらは文書論理構造を直接示すわけではないので領域を推定する場合は無視できる。また, 後に述べる C_{visual} となる画像がなかった場合, この強調タグでマークアップされた文字列が C_{visual} となる。

- 埋め込みタグ ($IMAGE, A, FORM, APPLET, OBJECT, EMBED, MAP$)

これらのタグは画像や, アンカといった特殊なオブジェクトを埋め込むために使用する。これらのタグの中に構造タグが入るということはない。本研究では, これらのタグは主に同期化可能なオブジェクトとしてカラーセルコンポーネントの取得時に使用する。

- その他のタグ

その他のタグはカラーセルコンポーネントを作成する場合重要ではないので無視する。

品川ら¹²⁾, Ashishら¹³⁾は本研究と同様にタグを分類して HTML 文書から論理木を導き出している。彼らは分類されたタグ構造を用途によりさらに分類しているのに対し, 本研究ではタグ構造の分類のみによる論理木の作成を行っている点が異なる。

以下に初期カラーセルのカラーセルコンポーネントの各構成部分の生成方法を述べる。

3.2 ヘッダ部分 C_{head} の生成

C_{head} は Web ページ P のタイトルである。つまり, P の $\langle title \rangle$ と $\langle /title \rangle$ に囲まれた領域が C_{head} となる。 P に $\langle title \rangle$ と $\langle /title \rangle$ がないとき, 下記に示す方法で抽出された音声部分の領域に見出しタグを含むならば, それをヘッダ部分とする。複数の見出しタグがある場合, トップレベルの方を1つ選ぶ。さらに音声部分の領域に見出しタグを含まないとき, C_{voice} の最初の数文字を C_{head} とする。

3.3 音声部分 C_{voice} の生成

C_{voice} は P の一部であり, ユーザの指定したすべての検索キーワード $K = \{k_1, k_2, \dots, k_n\}$ を含む文章または文である。 C_{voice} は音声読み上げを行う部分であり, C_{voice} のデータ量に応じて1つのカラーセルコンポーネントの呈示時間が決定される。Web-Carousel では, C_{voice} の文章長は, 原則として一定量 MAX_c を超えないように定める。

C_{voice} を取得する手順を図3と以下に示す。

- (1) P を構造タグで区切られた最小な領域に分け木構造を作成する。構造タグを用いることにより, C_{voice} の候補となる構成要素の文章が途中で不適切にタグにより切断されることはほとんどな

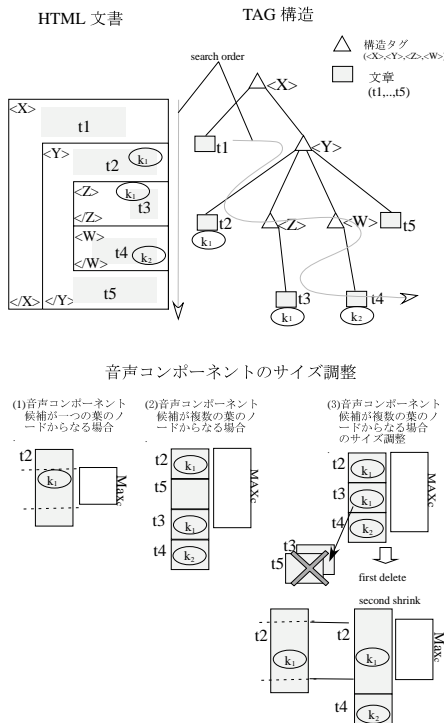


図3 カルセルコンポーネントの取得
Fig. 3 Discovery of Carousel Component.

く、それらの文章は図3の $t1, t2, t3, t4, t5$ のように葉節点となる。葉節点とは、子を持たない節点(ノード)のことを示す。

- (2) その木構造の先頭より横型探索を行い、少なくとも1つの検索キーワードが最初に出てくる葉節点を見つける。その最初に発見された葉節点を起節点とする。図3の場合 $t2$ が起節点となる。起節点にすべての検索キーワードが含まれていた場合、その時点で発見は終了する。
- (3) すべてのキーワードを発見するまで、木構造全体に関する横型探査を起節点からさらに継続し、次々と葉節点を C_{voice} の候補として加える。このようにして、すべての検索キーワードを含む C_{voice} の候補領域を求めることができる。たとえば図3の場合、 $t2$ を起節点として、 $t2, t5, t3, t4$ が C_{voice} の候補領域となる。
- (4) 上記で求めた C_{voice} の候補領域が、最大量 MAX_c を超えている場合、下記の手順で C_{voice} の容量の調整を行い、 C_{voice} を決定する。
 C_{voice} の候補領域が1つの葉節点のみで構成されている場合
 検索キーワードが最初に出現する文を発見し、その他の文を削除する。

C_{voice} の候補領域が複数の葉節点で構成されている場合

- まず検索キーワードが含まれていない葉節点を削除する。
 - 検索キーワードが含まれていない葉節点を削除しても MAX_c を超えている場合、同じ検索キーワードが再度出現した葉節点を検索方向とは逆方向から削除する。
 - 上記の作業を行った後、 C_{voice} の候補領域が MAX_c をまだ超えている場合、検索キーワードの出現順に各葉節点の文章から検索キーワードが含まれる文を MAX_c 分抽出する。つまり、図3の下部に示すように、 $t2$ がすでに MAX_c を超えている場合、 $k1$ を含む文を抽出し、 $t2$ の $k1$ を含まない文と $t4$ は削除する。
- (5) 上記(4)までで検索キーワードを1つのみ含む文を抽出した場合、それだけで長さが MAX_c を超えているケースもありうる。しかし、この場合は例外的に MAX_c を超えた文であっても C_{voice} として決定する。その文を MAX_c で切断すると、ユーザに文の意味が伝わらなくなる恐れがあるためである。

上記の方法で音声パート C_{voice} を決定する。

3.4 ビジュアルパート C_{visual} の生成

ビジュアルパート C_{visual} は C_{voice} と同期する1つの画像や動画である。 C_{voice} と同期する画像や動画がない場合、強調タグで囲まれた文字が C_{visual} となる。さらに強調タグで囲まれた文字がないとき、 C_{visual} は空となる。 C_{voice} の生成手順を下記に示す。

- C_{voice} と少なくとも1つの画像を含む極小のタグ領域を $R(C_{voice})$ とする。ここで、タグ $\langle X \rangle$ のタグ領域とは、構造タグ $\langle X \rangle$ を親とする部分木の、 $\langle X \rangle$ の子孫節点すべてからなる集合のことである。また、 C_{voice} が複数の葉節点で構成されている場合、 $R(C_{voice})$ の中にさらに小さいタグ領域が含まれていることがあるため、 $R(C_{voice})$ を極小のタグ領域と呼ぶ。Web ページに1つの画像もない場合、 $R(C_{voice})$ は C_{voice} とする。
- $R(C_{voice})$ の中に画像が1つの場合、その画像を C_{visual} とする。
- $R(C_{voice})$ の中に画像が複数存在する場合、 C_{voice} に最も近い画像を C_{visual} とする。 C_{voice} に最も近い画像が複数ある場合、木構造の左側にありかつ深さの浅い画像を C_{voice} とする。

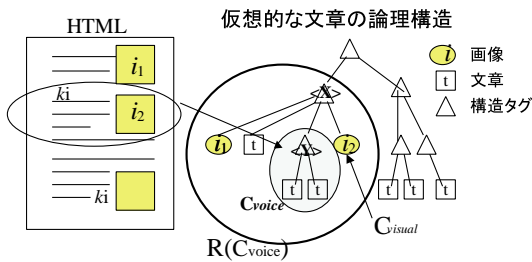


図4 C_{visual} 発見のイメージ図
Fig. 4 Discovery of C_{visual} .

- Web ページに 1 つの画像も含まない場合， $R(C_{voice})$ 中の強調タグで囲まれた文字が C_{visual} となる．
- 上記以外の場合， C_{visual} は空となる．このとき，画面に表示されるのは C_{head} のみである．

図4の場合， C_{voice} に最も近い画像は i_2 であるので， i_2 が C_{visual} となる．

上記の方法でカーセルコンポーネントを生成し，検索結果のランキング情報を元に，上位10件を初期カーセルとする．カーセルコンポーネントの C_{head} と C_{visual} を携帯端末の画面に表示し， C_{voice} を音声読み上げる．音声読み上げが終了した時点で次のカーセルコンポーネントを呈示する．このようにカーセル呈示を行う．

4. カーセルの動的再構成

4.1 意味的關係に基づいたカーセルの再構成

現在呈示しているカーセルの中にユーザがほしいカーセルコンポーネント (Web ページ) が存在しない場合，そのカーセルの中から1つの適当なカーセルコンポーネントを選択した後，簡単なインタラクションを行うことにより次のカーセルを取得できるようにする．

インタラクションとして以下の4つのカーセル生成要素を提案する．

- 類似カーセル生成：現在呈示しているカーセルコンポーネントに類似した情報を要求する場合
- 相違カーセル生成：現在呈示しているカーセルコンポーネントと異なった情報を要求する場合
- 詳細カーセル生成：現在呈示しているカーセルコンポーネントよりもっと詳細な情報を要求する場合
- 簡潔カーセル生成：現在呈示しているカーセルコンポーネントよりもっと簡潔な情報を要求する場合

WebCarousel はこれらユーザが指定したインタラ

クションの要素を取得し，動的にカーセルを再構成する．このとき，類似カーセル，相違カーセル，詳細カーセル，簡潔カーセルをどのように決めるかが問題になる．

Web ページの文書 P_i における特徴ベクトルを出現単語 W_1, \dots, W_n を各基底とする重要度 w_1^i, \dots, w_n^i として求めると，各文書の特徴ベクトル $F(P_i)$ は

$$F(P_i) = (w_1^i, \dots, w_n^i) = \frac{1}{N_i} (f_1^i, \dots, f_n^i)$$

で表される．ここで， $f_j^i, j \in \{1, \dots, n\}$ は Web ページの文書 P_i における単語 W_j の出現回数， N_i は P_i の総単語数を表す．すなわち w_j^i は単語の出現頻度 tf を表している．

2つの文書間についての特徴ベクトルを考え，以下のようなアルゴリズムを提案する．ここでは，選択されたカーセルコンポーネントである Web ページの文書 P_0 を基準として比較する文書 P_1 に関する類似度，相違度，詳細度，簡潔度をそれぞれ計算し， P_1 がどのタイプに最もあてはまるかを求め分類を行う．分類を行うにあたり，この4つの指標を下記のような式で定義した．また，その意味を直感的に理解させるために図示すると図5, 6, 7, 8, 9 のようになる．

- (1) 類似度の定義：類似度は比較する2つの文章間の類似の度合いを示す指標である．ここでは，比較する2つの文書を P_0, P_1 とし，コサイン相関値 $S(P_0, P_1)$ を用いて類似度 $\lambda_{similarity}$ を取得した．式で表すと下記のようになる．

$$\lambda_{similarity} = S(P_0, P_1) = \frac{F(P_0) \cdot F(P_1)}{\|F(P_0)\| \cdot \|F(P_1)\|}$$

$\lambda_{similarity}$ の値が高ければ P_1 は P_0 の類似文書となる．図6の場合， P_0 と P_1 の各単語の重要度がほぼ等しいため， $\lambda_{similarity}$ は高くなり，この2つの文書は類似している．

- (2) 相違度の定義：相違度は比較する2つの文章の相違の度合いを示す指標である．相違度は類似度の逆であり，2つの文書間の非類似度となる．相違度 $\lambda_{different}$ を式で表すと，上述のコサイン相関値より，

$$\lambda_{different} = 1 - S(P_0, P_1)$$

となる． $\lambda_{different}$ の値が高ければ，話題の違う文書である可能性が高く P_1 は P_0 の相違文書となる．図7の場合，基準文書の重要度が高い単語は比較文書での重要度が低く，逆に基準文書の重要度が低い単語は比較文書での重要度

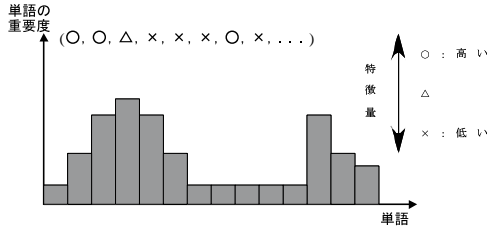


図5 基準となるページの特徴ベクトル

Fig. 5 The keyword vector of the standard page.

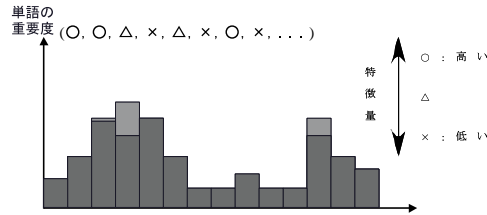


図6 類似ページの特徴ベクトル

Fig. 6 The keyword vector of the similar page.

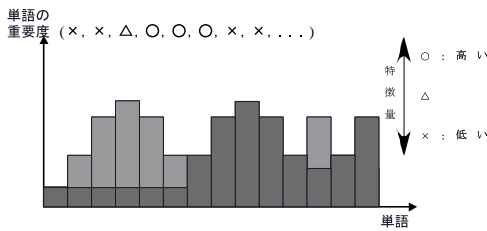


図7 相違ページの特徴ベクトル

Fig. 7 The keyword vector of the different page.

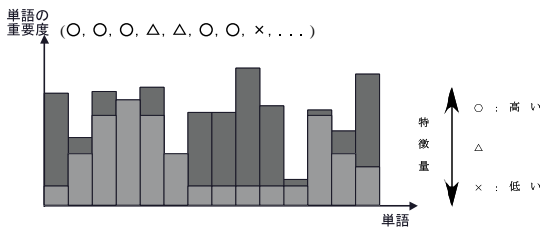


図8 詳細ページの特徴ベクトル

Fig. 8 The keyword vector of the more detailed page.

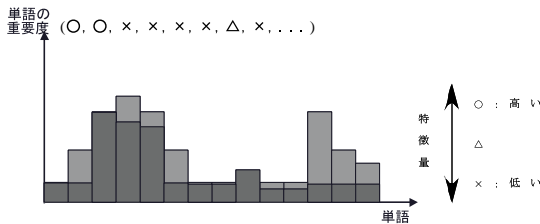


図9 簡潔ページの特徴ベクトル

Fig. 9 The keyword vector of the more summarized page.

が高くなっている．そのため， P_0, P_1 のコサイン相関値は低くなり， $\lambda_{different}$ は1に近く

なるため，2つの文書は相違している．

- (3) 詳細度の定義：詳細度は，比較文書が基準文書よりも多くの付加情報を持っている度合いを示す指針である．

詳細度を式で表す．まず，基準文書の特徴ベクトルを

$$F(P_0) = (w_1^0, \dots, w_i^0, w_{i+1}^0, \dots, w_j^0, \dots, w_n^0)$$

とし，比較文書の特徴ベクトルを

$$F(P_1) = (w_1^1, \dots, w_i^1, w_{i+1}^1, \dots, w_j^1, \dots, w_n^1)$$

とする．そのとき，詳細度は下記ようになる．

- $k \in \{1, \dots, i\}$,
 $w_k^0 > \theta_0, w_k^1 > \theta_0$ かつ,
 $|w_k^1 - w_k^0| \leq \theta_1$
 θ_0 は比較する単語を抽出するための単語の重要度の閾値を示し， θ_1 は重要度が類似している単語を削除するための閾値を示す．

- $k \in \{i+1, \dots, j\}$,

$$w_k^1 - w_k^0 \leq 0$$

のとき， w_k^0 と w_k^1 は0と見なす．

すなわち，上記条件を $F(P_0)$ と $F(P_1)$ に適用した $F'(P_0)$ と $F'(P_1)$ は次のようになる．

$$F'(P_0) = \underbrace{(0, \dots, 0, 0, \dots, 0)}_{1 \sim i} \underbrace{(0, \dots, 0)}_{(i+1) \sim j}, w_{j+1}^0, \dots, w_n^0$$

$$F'(P_1) = \underbrace{(0, \dots, 0, 0, \dots, 0)}_{1 \sim i} \underbrace{(0, \dots, 0)}_{(i+1) \sim j}, w_{j+1}^1, \dots, w_n^1$$

ここで，比較対象とならない単語を分かりやすく示すために， $F'(P_0)$ と $F'(P_1)$ の単語の並びを $1 \sim i$ は基準文書と比較文書の単語の重要度がともに大きくかつ基準文書と比較文書の重要度が類似している単語とし， $i+1 \sim j$ は比較文書の重要度が基準文書のそれより小さい単語の並びとした．

P_0 に対する P_1 の詳細度 λ_{detail} は下記のようになる．

$$\lambda_{detail} = \frac{1}{n-j} \sum_{k=j+1}^n (w_k^1 - w_k^0)$$

以上より， λ_{detail} が高い文書は注目すべきキーワードのほかに付加情報を多く持つより詳細な情報である．

図8，図10の場合，基準文書と比較文書における，重要度がともに小さい単語以外の単語の重要度が比較文書の方が大きく，詳細な文書となる．

- (4) 簡潔度の定義：簡潔度は，基準文書と比較文書

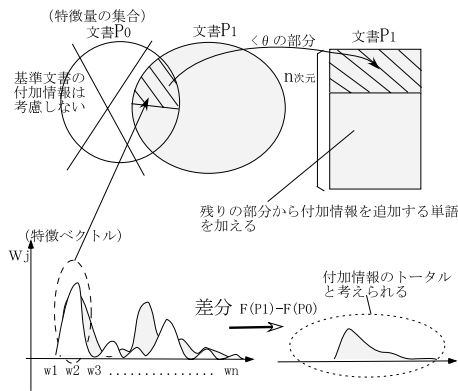


図 10 特徴ベクトルをもとにした 2 文書間の詳細関係

Fig. 10 Detailed relation between two documents based on the weight of element.

が類似し、かつ比較文書が基準文書より情報量が少ない度合いを示す指針である。つまりは、基準文書と比較文書における、重要度がともに小さい単語以外の単語の重要度が比較文書の方が小さい文書が簡潔な文書となる。よって、簡潔度は詳細度の逆である。簡潔度を式で表すと、下記ようになる。

- $k \in \{1, \dots, i\}$,
 $w_k^0 > \theta_0, w_k^1 > \theta_0$ かつ,
 $|w_k^0 - w_k^1| \leq \theta_1$
- $k \in \{i+1, \dots, j\}$,
 $w_k^0 - w_k^1 \leq 0$

この場合 w_k^0 と w_k^1 は 0 と見なす。したがって、 P_0 に対する P_1 の簡潔度 λ_{summ} は下記ようになる。

$$\lambda_{summ} = \frac{1}{n-j} \sum_{k=j+1}^n (w_k^0 - w_k^1)$$

以上より、 λ_{summ} が高い文書は注目すべきキーワードのほかに付加する情報が基準文書に比べ少く、より簡潔な情報として位置付ける。

図 9 の場合、重要度がともに小さい単語以外の単語は比較文書が基準文書より重要度が低いため、比較文書は簡潔な文書となる。

4.2 カラーセルコンポーネントの再生成

カラーセルを動的に再構成した後、そのカラーセルを構成するカラーセルコンポーネントを作成する。

カラーセルの再構成時に基準文書 P と比較文書 P' を比較することで、2 つの文書間でともに重要度の高い単語を検出できる。この単語を関係キーワード K' と呼ぶ。 K' は検索キーワード $K = \{k_1, k_2, \dots, k_n\}$ と同様に重要であると考え、カラーセルコンポーネン

トの再生成時に使用する。 P の特徴ベクトルを $F(P)$ 、 P' の特徴ベクトルを $F(P')$ とし、 $F(P)$ を構成する各々の単語の重要度 w_i の平均値を閾値 θ とする。そのとき、類似カラーセル、詳細カラーセル、簡潔カラーセルの K' は、単語の重要度が $F(P)$ 、 $F(P')$ において θ より大きい単語である。相違カラーセルにおいては、相違度を決定する方法の性質上、そのような K' は存在しないので無視する。

類似カラーセル、詳細カラーセル、簡潔カラーセルの再構成時の各カラーセルコンポーネントの C_{voice} は、初期カラーセルの C_{voice} と同様の方法で決定するが、検索キーワード K と関係キーワード K' を含む領域となる。 C_{voice} の候補領域がカラーセルコンポーネントの呈示制限 MAX_c を超えた場合、下記の優先順位で C_{voice} を決定する。

- (1) K と K' の両方が含まれる領域。
- (2) K を含むが K' を含まない領域。
- (3) K' を含むが K を含まない領域。

相違カラーセルは関係キーワード K' は存在しないため、カラーセル再構成時の C_{voice} は、初期カラーセルの C_{voice} と同様の方法で決定する。

5. プロトタイプシステム

5.1 概要

実際にプロトタイプシステムを、携帯電話端末のエミュレータを用いて、PC 上で実装した。

本システムはすべて Windows プラットフォーム上で作成した。主な開発環境として Microsoft VisualC++ を使用した。また、検索エンジンのインデクスエンジン機能に、オラクル社の DBMS Oracle8i 8.1.5 Workgroup Server を利用した。

プロトタイプシステムの概要は以下のとおりである (図 11 参照)。

- (1) ユーザからの問合せにより、検索エンジンを用いて Web の検索を行う。今回は自作の検索エンジンを用いた。
- (2) 得られた検索結果の Web ページを取得し、上位 10 件の Web ページから最初のカラーセルコンポーネントを作成する。
- (3) (2) で作成した 10 件のカラーセルコンポーネントで初期カラーセルを作成する。
- (4) 初期カラーセルのうちの 1 つのカラーセルコンポーネントを呈示する。
- (5) ユーザからのインタラクションがない場合、システムは次のカラーセルコンポーネントを呈示する。

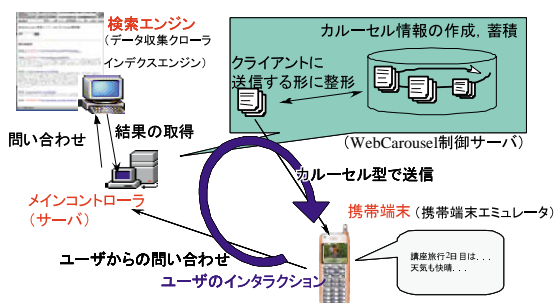


図 11 プロトタイプシステム概要

Fig. 11 The prototype system architecture.

- (6) ユーザからのインタラクションがあった場合、システムはユーザから指定された類似カルーセル、相違カルーセル、詳細カルーセル、簡潔カルーセルのいずれかを再構成する
- (7) カルーセルの再構成後、関係キーワードを含めたカルーセルコンポーネントを再生成する。
- (8) 生成されたカルーセルコンポーネントを順次表示する。

クライアント側では送られてきたカルーセルコンポーネントについて、ヘッダパートとビジュアルパートを画面上で表示しながら音声読み上げ機能により音声パートを読み上げる。音声読み上げ機能には Microsoft Speech API を用いた。実際にエミュレータで閲覧する場合、サイズの大きな画像は縮小してクライアントに送る。その後、4つの簡単なインタラクションをユーザが行うことによって次に表示するカルーセルコンポーネントを動的に再構成する。

5.2 プロトタイプシステムの詳細

プロトタイプシステムは以下の3つの部分から構成される。

- (1) WebCarousel 検索エンジン
- (2) WebCarousel 制御サーバ
- (3) 携帯端末エミュレータ

5.2.1 WebCarousel 検索エンジン

WebCarousel 検索エンジンはデータ収集クローラ機能とインデクスエンジン機能の2つの機能からなる。データ収集クローラは、入力されたシード URL とインデクス作成サイトの指定から、一般の検索ロボットと同様にリンク情報をたどり、そのサイトのページを収集する。

インデクスエンジン機能は、文書の特徴ベクトルおよび、文書間の関係情報をあらかじめ DBMS に保管しておくことで、実際の動作スピードを向上させる。WebCarousel では動的なグループ化を行うため、複数のユーザからアクセスがあった場合のサーバの処理

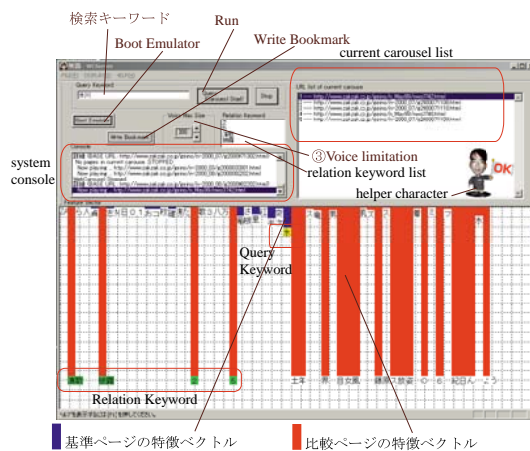


図 12 WebCarousel 制御サーバ

Fig. 12 Display of WebCarouel control server.

スピードが問題となる。そのため、あらかじめ文書間の関係情報を保存しておくことで高速な再構成を実現するようにしている。インデクスエンジンには、以下の情報を保存する。

- 文書の解析に利用したキーワード群の情報
- 文書各々の特徴ベクトル情報
- 文書間の関連情報(類似, 相違, 詳細, 簡潔)と、関連キーワード情報

5.2.2 WebCarousel 制御サーバおよび携帯端末エミュレータ

本プロトタイプシステムの中核部分であり、ユーザからの問合せを基に検索結果を参照して、動的にカルーセルを構成し、クライアント端末へカルーセルを送信する。

図 12 に制御サーバの画面を、図 13 に携帯端末エミュレータの画面を示す。

5.3 プロトタイプシステムの評価実験

本プロトタイプシステムの実験は、WebCarousel 制御サーバの精度実験と携帯端末エミュレータの評価実験を行った。

5.3.1 WebCarousel 制御サーバの精度実験

4.1 節で述べた手法を用いて2つの実験を行った。本実験は検索者が欲しい情報を取得できたかどうかの主観的判断を用い、今回は被験者1人で評価した。

検索結果の件数による適合率の実験

被験者は下記の実験を異なる検索キーワードを用いて10回行った。1回の検索で約600~1700件の検索結果を取得した。

- (1) 被験者は任意の検索キーワードを入力し、WebCarousel 検索エンジンにより検索結果を取得する。

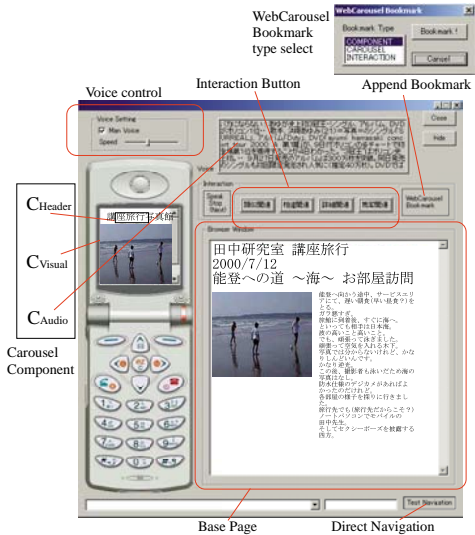


図 13 プロトタイプシステムエミュレータ (クライアント)
Fig. 13 Display of a prototype system emulator.

表 1 実験結果 (各分類の正解の出現率 (適合率))
Table 1 Result of the experiment.

| 検索結果数 | 類似文書 | 相違文書 | 詳細文書 | 簡潔文書 |
|-------|------|------|------|------|
| 100 件 | 90% | 89% | 73% | 30% |
| 200 件 | 82% | 85% | 70% | 30% |
| 300 件 | 82% | 80% | 59% | 21% |
| 500 件 | 73% | 75% | 45% | 17% |

- (2) 得られた検索結果から上位 100 件, 200 件, 300 件, 500 件を抽出し, これらを評価対象の検索結果として各件数ごとに (3) ~ (5) を行う.
- (3) 評価対象の検索結果から, 被験者は基準となる Web ページ (文書) を任意に選ぶ.
- (4) 選んだ基準となる Web ページに対し, 類似カーセル生成, 相違カーセル生成, 詳細カーセル生成, 簡潔カーセル生成の 4 つのインタラクションを行い, 得られた各々の結果, 上位 10 件をシステムの解とする.
- (5) 4 つのインタラクションごとに得られたシステムの解の Web ページと, 基準となる Web ページ (文書) を目視により比較し, その適合率を求めた.

表 1 に実験結果を示す.

類似カーセル, 相違カーセル, 詳細カーセル, 簡潔カーセルすべてにおいて, 評価対象の検索結果数が増えるにつれて適合率が減少しているが, 類似カーセル, 相違カーセルについては本実験結果の適合率が平均 80% を超える高い適合率が得られた. また, 詳細カーセルは評価対象の検索結果数が 200 件

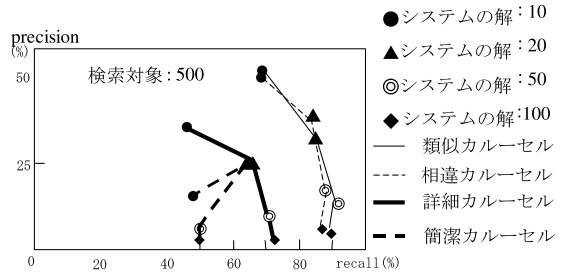


図 14 適合率と再現率
Fig. 14 Precision and recall.

までは 70% の適合率であるため, 検索結果数が少ない場合はある程度有効であると思われる. しかし, 検索結果が多い場合, 適合率が下がり, 相違する文章になってしまうという結果が得られた. 簡潔カーセルについては適合率が 30% 以下と低い結果となった.

再現率の実験

- (1) 検索対象の全 Web ページを 500 件とし上記適合率の実験を行う. このとき, 検索結果は 10 件, 20 件, 50 件, 100 件の 4 とおりとする.
- (2) この検索対象の Web ページ 500 件を, 目視により基準となるページに対する類似カーセル, 相違カーセル, 詳細カーセル, 簡潔カーセルに分類する.
- (3) (1) のシステムの解と (2) の結果から再現率を求めた.

図 14 にこの実験結果を示す.

図 14 では, 類似カーセル, 相違カーセル, 詳細カーセルにおいては評価対象の検索結果数が増えるにつれて適合率が減少し, 再現率が増加しているのが分かる. 簡潔においては, 検索結果数と適合率, 再現率の関係が様ではないが, 全般的に, 適合率再現率とも低い値を示しているため, 本実験の結果の手法の再検討が必要であることが分かる.

上記 2 つの実験により, 類似カーセル, 相違カーセル, 詳細カーセルにおいては適合率, 再現率ともある程度高い値を示し, 十分な精度が得られたと考えられる. しかし, 簡潔カーセルにおいては, 適合率再現率とも低く, 本手法の改良が課題となった. 本実験で用いた単語の重要度から類似カーセル, 相違カーセル, 詳細カーセル, 簡潔カーセルに分類する手法は数ある手法のうちの 1 つであり, 様々な分類の手法が考えられる. 平田ら¹⁴⁾はその文書に出現する単語量と, 単語の共起関係を利用してマッピングを行い, これらの分類を行う提案とその実験を行っている. これらの分類をより効果的に行うためには様々な点を改良し実験を重ねていく必要がある.

表 2 カラーセルコンポーネントの評価実験結果 単位(人)
Table 2 Result of the evaluation experiment for Carousel Components.

| 質問事項 | はい | いいえ | どちらでもない |
|--|----|-----|---------|
| カラーセルコンポーネントの音声は聞き取りやすいか | 6 | 0 | 2 |
| カラーセルコンポーネントの音声情報は検索結果のページの重要な部分を示しているか | 8 | 0 | 0 |
| カラーセルコンポーネントの画像は音声情報と同期しているか | 8 | 0 | 0 |
| カラーセルコンポーネントのタイトルの呈示は分かりやすいか | 6 | 0 | 2 |
| カラーセルコンポーネントは Web ページの内容を把握するために有効な方法である | 7 | 0 | 1 |

5.3.2 携帯端末エミュレータの評価

携帯端末エミュレータの評価実験ではカラーセルコンポーネントの評価と WebCarousel 全体の評価の 2 つの実験を行った。被験者は通常、携帯電話の一般の検索サービスを使用している研究室の学生 8 人である。被験者はプロトシステムを実際に操作してアンケートに答えてもらった。

カラーセルコンポーネントの評価実験

カラーセルコンポーネントの有用性を計るために、8 人の被験者に対し、スポーツ・芸能ニュースより「松井」という同じキーワードを与え、検索を行った。検索結果の初期カラーセルを構成する 10 個のカラーセルコンポーネントとその実際のページを比較してもらった。アンケートの結果を表 2 に示す。

- 音声の聞き取りやすさの回答では、2 人が「どちらともいえない」という回答であるが、その理由は、抑揚がない、漢字の読みが誤っているなどの音声読み上げソフトに関する不満であった。
- 音声部分の抽出や画像の抽出に関しては、8 人全員が満足していた。これは、今回の実験は画像が多く、またその画像が何を指し示しているのかが分かりやすいスポーツ・芸能ニュースを対象としたためと考えられる。
- タイトルの呈示に関しては、「タイトルが長い場合、画面のほとんどをタイトルの文字列で占めてしまうため、画像が見えない」という不満であった。
- カラーセルコンポーネントは Web ページの内容を把握するために有用な方法であるかの質問では、「移動中など検索結果を片手間に見るときに効果的である」、「音声読み上げだけでなくタイトルと画像表示があるため、情報がより分かりやすくなっている」という理由から有効であると答えた回答に対し、「画像がない場合、欲しい情報かどうか把握できないページもある」という回答があった。

また、カラーセルコンポーネントの内容が理解できた比率は、90%が 1 人、80%が 3 人、70%が 2 人、60%が 2 人という結果であり、平均が 70%を上回る結

果となった。

上記の結果より、カラーセルコンポーネントは Web ページの内容を把握するためにほぼ有用であると考えられるが下記の 2 点が課題となった。

- ヘッド部分のタイトルの文字数の制御の検討
- 画像がない Web ページの場合のビジュアルパートの検討

WebCarousel 全体の評価実験

先の 8 人に、WebCarousel の一連の操作をしてもらい、通常使用している携帯電話での一般の検索サービスと比較する実験を行った。実験データはスポーツ・芸能ニュースを用いた。そして、被験者は各自好きな検索キーワードを 5 種類入力し、各々の検索キーワードに対して類似カラーセル生成、相違カラーセル生成、詳細カラーセル生成、簡潔カラーセル生成の 4 つのインタラクションすべてを行った。

表 3 に評価実験結果を示す。

評価実験の結果より下記のことが判明した。

- カラーセルの呈示方法は携帯電話での検索結果の閲覧方式として有効であると考えた人が 8 人中 6 人 (75%) であった。有効である理由としては、「各ページごとにインタラクションを行わなくても自動的に検索結果が呈示されるため有効である」、「見逃したカラーセルは待っていればまた呈示されるため、ぼーっとして見るときに便利である」などの回答があった。一方、どちらでもない理由としては、「急いで検索するとき、検索結果のページをいちいち見ていなければならないため、時間がかかってしまう」などの回答があった。
- 欲しい検索結果を取得できた人は 8 人中 6 人 (75%) であった。検索結果を取得できた人に共通する点は、入力した検索キーワードとして、複数の別の意味を持つ単語を入力した点であった。たとえば「ヒット」という単語は、野球のヒットを示す場合とヒット曲の「ヒット」を示す場合では、まったく異なる意味を持っている。このとき、類似カラーセル生成、相違カラーセル生成、詳細

表3 WebCarousel 全体の評価実験結果 単位(人)
Table 3 Result of the evaluation experiment for WebCarousel.

| 質問事項 | はい | いいえ | どちらでもない |
|---|----|-----|---------|
| カルーセルの呈示方法は携帯電話での検索結果の閲覧方法として有効である | 6 | 0 | 2 |
| 欲しい検索結果を取得できた | 6 | 1 | 1 |
| 類似(相違, 詳細, 簡潔)カルーセル生成は分かりやすいか | 5 | 0 | 3 |
| 類似(相違, 詳細, 簡潔)カルーセル生成は使いやすいか | 6 | 0 | 2 |
| 類似(相違, 詳細, 簡潔)カルーセルの分類に満足である | 2 | 0 | 6 |
| 類似(相違, 詳細, 簡潔)カルーセル生成のインタラクションのあとからカルーセル呈示されるまでの待ち時間が長い | 8 | 0 | 0 |

カルーセル生成, 簡潔カルーセル生成のインタラクションを用いて, 取得したい検索結果を得ることができた. 一方, 検索キーワードとして「巨人軍」などの固有かつある集合を示すような検索キーワードを入力したとき, 取得したい情報が十分に取り出せないことが分かった.

- 類似カルーセル生成, 相違カルーセル生成, 詳細カルーセル生成, 簡潔カルーセル生成のインタラクションについては, 分かりやすさ, 使いやすさはある程度良い結果となった. しかし, それらの分類の満足度は, 8人中6人が「どちらともいえない」であり, 6人とも「簡潔カルーセルに, 基準となるページの簡潔ではないページがあった」と簡潔カルーセルの結果に関する不満があることが分かった. これは, 5.3.1 項の実験結果と同様に, 簡潔カルーセルの手法の再検討が必要であることを示している.
- 類似カルーセル生成, 相違カルーセル生成, 詳細カルーセル生成, 簡潔カルーセル生成のインタラクション後から次のカルーセル呈示までにかかる時間については全員が長いと感じている. 実際にかかった時間は, 検索キーワードにより, 7分程度かかるものもあり, リアルタイム性にかけるものもあった.

以上の実験結果より, WebCarousel は携帯端末の検索手法として有用であると思われるが, 下記の課題があげられる.

- ユーザの検索時間にかけることのできる時間による, カルーセルのスピード調整
- キーワードの種類による適合率の実験
- 簡潔カルーセルの手法の再検討
- カルーセルの動的再構成時の解析速度の最適化

また, 今回の実装では, カルーセルを1方向のみで呈示したため, 見落としたカルーセルを再度見たいときは, カルーセルが一巡するのを待たなければならない. ゆえに, カルーセルの逆回転機能を付け加える必

要がある.

6. おわりに

本論文では, 携帯端末における Web ページの検索結果の受動的呈示方法として WebCarousel を提唱した. 携帯端末は, 通信速度, 表示領域, インタラクションの制限があるが, WebCarousel を用いることにより, これらの制限を解決して Web ページの検索結果を呈示することが可能となる. 本論文で述べた我々の研究を下記に示す.

- 携帯端末においての Web ページの検索結果の受動的視聴
 - Web ページの一部の音声読み上げと画像呈示
 - カルーセル呈示による自動反復呈示
 - 簡単なインタラクションによる膨大な検索結果の呈示
- Web ページの意味的關係に基づいたカルーセルの動的再構築

WebCarousel の特徴は

- カルーセル呈示のため, 同じコンテンツを何度でも見ることが可能である.
- 画像と音声読み上げからなる受動的呈示のため, 容易に片手間に検索結果を取得することができる.
- ユーザは類似カルーセル生成, 相違カルーセル生成, 詳細カルーセル生成, 簡潔カルーセル生成の簡単なインタラクションにより, 検索結果を取得することができる.

である.

本論文では, 単一のページを検索結果の対象とするカルーセルコンポーネントを定義した. しかし, 検索結果の集合をさらにカルーセルの単位として考えることも可能である. また本論文で用いた単語の重要度から類似カルーセル, 相違カルーセル, 詳細カルーセル, 簡潔カルーセルに分類する手法は一番手始めのものであり, 今後, 様々な点を改良し実験を重ねていく必要がある.

謝辞 本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501)および(株)NTTドコモとの共同研究による。ここに記して謝意を表す。

参考文献

- 1) WAP Forum のホームページ
http://www.wapforum.org/
- 2) 平川泰之, 遠山緑生, 安田絹子, 石川雅康, 浅田卓哉, 服部隆志, 萩野達也: 次世代モバイル Web システム, 産学官交流シンポジウム YRP 移動体通信産学官交流シンポジウム (2000).
- 3) Bukukkokten, O., Garcia-Molina, H. and Paepcke, A.: Focused Web Searching with PDAs, *9th International World Wide Web Conference*, Amsterdam (May 2000).
- 4) Hori, M., Kondoh, G., Ono, K., Hirose, S. and Singhal, S.: Annotation-Based Web Content Transcoding, *9th International World Wide Web Conference*, Amsterdam (May 2000).
- 5) Maly, K. and Zubair, M.: Personalized Portal for Wireless Devices, *10th International World Wide Web Conference*, Hong Kong, pp.74-75 (May 2001).
- 6) Rocchio, J.: Relevance Feedback in Information Retrieval, *The Smart System Experiments in Automatic Document Processing*, pp.313-323, Prentice Hall, Englewood Cliffs, NJ (1971).
- 7) 近藤宏行, 角谷和俊, 田中克己: 番組メタファアを用いた情報検索結果の呈示方式, 情報処理学会研究報告, 99-DBS-119-70, pp.419-424 (1999).
- 8) 服部多栄子, 角谷和俊, 灘本明代, 草原真知子, 田中克己: 番組メタファアによる Web ページの利用者適応型呈示方式, 情報処理学会研究報告, 99-DBS-119-69, pp.413-418 (1999).
- 9) 灘本明代, 服部多栄子, 近藤宏行, 沢中郁夫, 草原真知子, 田中克己: Web 情報の番組化のためのオーサリング機構, 情報処理学会研究報告, 00-DBS-120-14, pp.99-106 (2000).
- 10) 服部多栄子, 沢中郁夫, 灘本明代, 田中克己: Web の受動的視聴のための同期化可能領域の発見と番組化用マークアップ言語 S-XML, 情報処理学会研究報告, 00-DBS-121-2, pp.9-16 (2000).
- 11) 灘本明代, 服部多栄子, 近藤宏行, 沢中郁夫, 田中克己: Web コンテンツの受動的視聴のための自動変換とスクリプト作成マークアップ言語, 情報処理学会論文誌: データベース (TOD8), pp.103-116 (2001).
- 12) 品川徳秀, 北川博之: ユーザ視点に即した仮想 WWW ページの動的生成による閲覧支援, 情報処理学会研究報告, 99-DBS-119-71, pp.425-430

(1999).

- 13) Ashish, N. and Knoblock, C.A.: Wrapper Generation for Semi-Structured Internet Sources, *ACM SIGMOD Records*, Vol.26, No.4, pp.8-15 (1999).
- 14) 平田陽一, 松倉健志, 田島敬史, 田中克己: Web 検索における意味的適合フィードバック機構, 情報処理学会研究報告, 00-DBS-122-21 (2000).
- 15) Anupam, V., Freire, J., Kumar, B. and Lieuwen, D.: Automating Web Navigation with WebVCR, *9th International World Wide Web Conference*, Amsterdam (May 2000).

(平成 13 年 4 月 7 日受付)

(平成 13 年 6 月 28 日採録)

(担当編集委員 仲尾 由雄)



灘本 明代 (正会員)

1987 年東京理科大学理工学部電気工学科卒業, 同年 (株) TL ヤマギワ研究所入社 (株) 計算流体力学研究所, クボタシステム開発 (株) を経て 1995 年 (株) 関西新技術研究所入社。ソフトウェア開発, 可視化技術の研究開発に従事。1998 年より神戸大学大学院自然科学研究科情報メディア科学専攻博士課程に在籍。



近藤 宏行

1999 年神戸大学工学部情報知能工学卒業, 2001 年神戸大学大学院自然科学研究科修士課程修了, 2001 年 NHK 入社, NHK 徳島放送局勤務。現在に至る。データベース, ハイパーメディアに興味を持つ。



田中 克己(正会員)

1974年京都大学工学部情報工学科卒業，1976年同大学大学院修士課程修了．1979年神戸大学教養部助手，1986年同大学工学部助教授．1994年同教授(情報知能工学科)．

1995年同大学大学院自然科学研究科情報メディア科学専攻専任教授，2001年京都大学大学院情報学研究科社会情報学専攻教授，現在に至る．工学博士．主にデータベースの研究に従事．人工知能学会，日本ソフトウェア科学会，IEEE Computer Society，ACM等各会員．
