

# 閲覧履歴を反映したコンテキスト依存型 Web ブックマーク

中島伸介<sup>†</sup> 黒田慎介<sup>†</sup> 田中克己<sup>†</sup>

従来の Web ブックマークは、ユーザが興味を持った Web ページの URL とタイトルを保持しているのみである。しかしながら、ブックマークまでのプロセス(コンテキスト)には、ユーザの意図やブックマーク行為の価値が隠されていると考えられる。そこで、ユーザの閲覧履歴を反映した Web ブックマーク、すなわち“コンテキスト依存型 Web ブックマーク”を提案する。このコンテキスト依存型 Web ブックマークは、閲覧した Web ページとブックマークを付けた Web ページを比較することにより、Web ブックマークの代表的キーワードやランキング値をメタデータとして保持させるものである。本論文では、コンテキストデータの抽出方法を提案するとともに、コンテキスト依存型 Web ブックマーク機能を有する Web ブラウザのプロトタイプを実装し、その有用性について検証する。

## A Context-dependent Web Bookmark Reflecting Browsing Histories

SHINSUKE NAKAJIMA,<sup>†</sup> SHINSUKE KURODA<sup>†</sup>, and KATSUMI TANAKA<sup>†</sup>

The conventional Web bookmarks have only the URLs and the titles of Web pages of a user's interest. However, we believe that the context (the process of bookmarking) contains the user's intention and the value of the bookmarks. Therefore, we propose a context-dependent Web bookmark mechanism that reflects user's browsing history. The context-dependent Web bookmarks have meta data: the representative keywords and the ranking value of the bookmark computed by comparing the bookmarked Web page with the other pages in the context. We describe the usage of the context-dependent bookmarks. We also evaluate the context-dependent Web bookmark mechanism by using a prototype of Web browser with the mechanism.

### 1. はじめに

人間の活動によって生み出される成果に比べ、その成果に至るプロセス情報が記録・保存されることは少ない。しかしながら、そのプロセスにこそ、作業上のノウハウや、成果に対する人間の目的・意図が隠されていることが多い。したがって、人間の作業プロセスを取得して、有効利用することの意義は大きいといえる。また、Web 利用時の成果に対しても、そのプロセスを抽出・解析して、利用することの意義は大きいと考えられる。

Web 情報検索の成果品の 1 つは、ブックマークであると考えられる。しかし、従来のブックマークが保持するデータは、URL とタイトルのみであり、その価値や意味が不明確で分かりにくい。そこで著者

らはこれまで、ブックマークまでのプロセスを解析することにより、ブックマーク行為の価値や意味の表現について検討してきた<sup>1);2)</sup>。ブックマーク行為の価値とは、Web 閲覧時のブックマークまでのプロセスにおいて、量的、質的に、ユーザがどれほど詳しく調べたのかということの評価したものと考えられる。また、ブックマークの意味とは、ブックマークに対するユーザの意図であると考えられる。これは、ユーザが閲覧した Web ページの中で、必要であった Web ページ(ブックマークされたページ)と不要であった Web ページ(ブックマークされなかったページ)の特徴の違いにより、表現できるものと考えられる。

本論文では、ブックマークまでのプロセス(コンテキスト)を取得し、解析することにより、ブックマーク行為の価値や意味を表現することが可能な、“コンテキスト依存型 Web ブックマーク”を提案する。

なお、本論文では、ブックマークに関連する Web ページの閲覧履歴を“コンテキスト”と呼び、その閲覧範囲を“コンテキスト範囲”と呼ぶ。また、このコンテキストから抽出されるブックマークのメタデータ

<sup>†</sup> 京都大学大学院情報学研究科社会情報学専攻  
Department of Social Informatics, Graduate School of Informatics, Kyoto University  
現在、株式会社エヌ・ティ・ティ・ドコモ  
Presently with NTT DoCoMo, Inc.

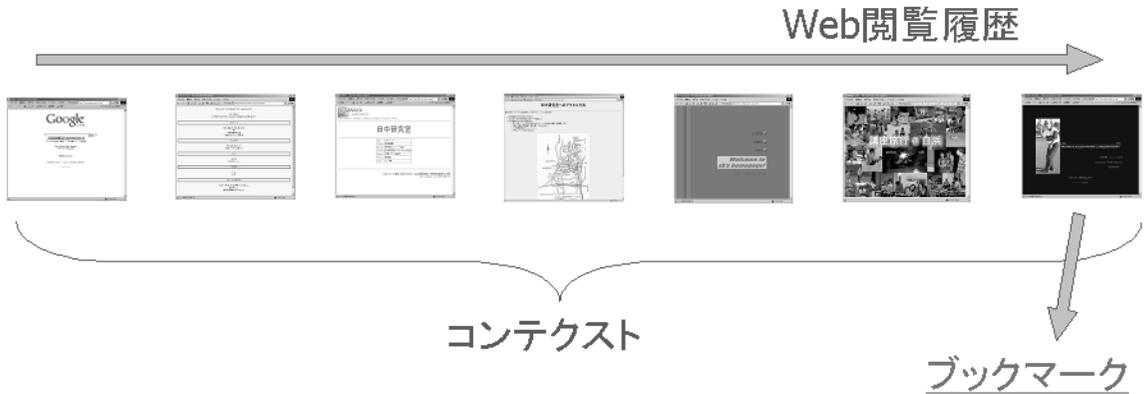


図1 Webブックマークのコンテキスト  
Fig. 1 Context of Web bookmark.

を“コンテキストデータ”と呼ぶ。

以上より、コンテキスト依存型 Web ブックマーク (以下、コンテキストブックマークと略す) の定義は、「Web ページの閲覧履歴であるコンテキストと、そこから抽出されるコンテキストデータをメタデータとして保持したブックマーク」とする(図1参照)。

すなわち、このコンテキストブックマークの有用性について検討することが、本論文の目的である。

本論文の成果の概要は以下のとおりである：

- ブックマーク行為の価値を示す、ブックマークランキング値の算出方法について検討した。
- ブックマークの意味、すなわちブックマークに対するユーザの意図を表現するための方法として、ブックマークした Web ページおよび閲覧した他の Web ページの特徴抽出方法について検討した。
- コンテキストブックマークの利用用途の提案および、評価実験に基づいた有用性を検証した。
- ブックマーク時のユーザの行動モデルに基づいた、各ブックマークに係る閲覧範囲(コンテキスト範囲)の限定方法について検討した。

以下、本論文の構成を示す。2章では関連研究について述べる。3章ではコンテキストブックマークのメタデータ(コンテキストデータ)の具体的な抽出方法について述べる。4章ではプロトタイプシステムの実装について述べる。5章ではコンテキストブックマークの利用用途について述べる。6章ではコンテキストブックマークの有用性の評価について述べる。7章ではブックマーク時のユーザの行動モデルを考慮したコンテキスト範囲の限定方法について述べる。8章では結論と今後の課題について述べる。

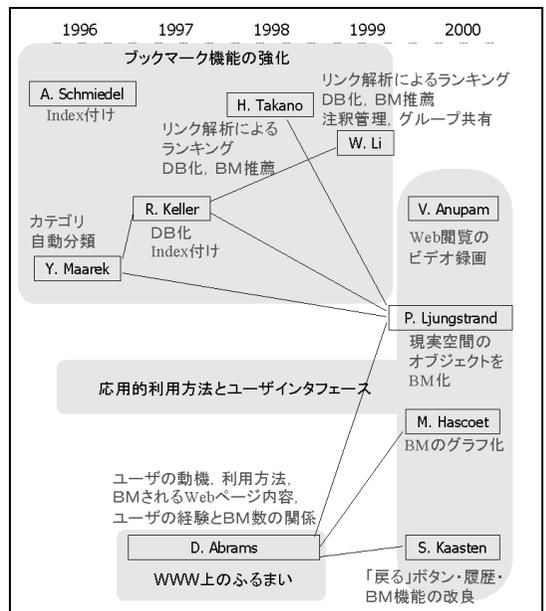


図2 関連研究の分類と特徴  
Fig. 2 Related works.

## 2. 関連研究

ブックマークに関連した研究は、数多く行われている。調査したもののうち、主な研究を図2に示す。各研究を結ぶ直線は引用関係を示している。

図2に示すとおり、ブックマークに関する研究は、  
1) ブックマーク管理の効率化・高度化に関する研究、  
2) 応用的利用やユーザインタフェースに関する研究、  
3) ユーザの振舞いに関する調査研究、の3つに分類することができる。

## 2.1 ブックマーク管理の効率化・高度化に関する研究

ブックマークは大量の URL データであることと、タイトルや URL のみで情報が十分でないことから、分類や組織化を手動で行うのは面倒である。これに対して、ブックマークの管理や組織化を支援するようなシステムが提案されている。

Maarek らは、ユーザが新たに追加しようとするブックマークの自動カテゴリ分類を行うことができる、Bookmark Organizer を提案している<sup>3)</sup>。このシステムでは、階層化されたカテゴリに対してシステムがブックマークの自動分類を行うことができる。

Schmiedel らは、ブックマークの管理方法として従来行われている、キーワードやテキスト文による注釈付け、フォルダへの分類によるカテゴリ分類等の各種手法の問題点を指摘している。この問題点に対し、本の目次のようなトピックの階層構造により、その特徴を表現する手法を提案している<sup>4)</sup>。このシステムにより、ブックマークが示す Web ページの複雑な意味の内容を表現することを可能にしている。

Keller らは、単純に階層化したフォルダにブックマークを分類するのみでは不十分であると主張し、ブックマークの管理・組織化・共有のための新しいシステム (WebTagger) を提案している<sup>5)</sup>。URL が示す Web ページは、多くの側面を持つために、1 つのカテゴリフォルダに分類することはできない。また、複数のコピーを作成して、複数のフォルダに格納することも、煩雑になり管理が面倒になる。したがって、各々のブックマークに、属する複数のカテゴリ名等のメタデータを保持させ、ブックマークをデータベース化することにより、ブックマークを管理する。このシステムでは、ブックマークの検索・共有・ランキングを可能にしている。

Takano らの Dynamic Bookmark<sup>6)</sup>では、ユーザが閲覧した Web ページ集合を解析し、訪れたページとそこから巡航可能な Web ページを、リンク構造 (Authority と Hub) と再訪回数から計算して、新たなブックマーク候補として、ランク値とともにユーザに提示することができる。

Li らの Power Bookmark<sup>7)</sup>では、ブックマーク管理の効率化に関する多くの機能を提供している。特に、ブックマークへの注釈管理を行うことにより、グループ内でのブックマーク共有を可能にしている。

これらの研究は、ブックマークを効率的に利用できるように仕組みの提案という意味では、研究の目的に共通点がある。しかしながら、コンテキストデータの

解析に基づくブックマークの価値や意味を算出し、利用するようなアプローチはない。

## 2.2 応用的利用やユーザインタフェースに関する研究

Ljungstrand らの WebSticker<sup>8)</sup>では、仮想空間と現実空間のオブジェクトを結び付けることを目的として、現実空間のバーコードを Web 情報に対するブックマークとして利用することができる。あらかじめバーコードを印刷した Post it やバーコード付きのオブジェクトを Web ページと関連付けることにより、現実空間のオブジェクトをブックマークとして扱うことができる。同じバーコードを持つオブジェクトを複数人数で保持することで、ブックマークの共有も可能にしている。

Hascoet は、最近閲覧した Web ページ、最も多く閲覧した Web ページ、すべては読み切っていないが後で見ようとしている Web ページ、そしてすべてのブックマークとそれぞれの関係を表したグラフを示すことにより、ユーザのブックマーク管理支援を行う、Bookmap というシステムを提案している<sup>9)</sup>。

Kaasten らは、「閲覧される Web ページの 60% が以前に訪れたことのあるページである」というデータから、ブラウザの「戻る」ボタン、履歴、ブックマーク機能の重要性を提唱するとともに、これらの機能の問題点を解決するシステムを提案している<sup>10)</sup>。

Anupam らは、単一の Web ページではなく、あたかもビデオ録画するように、ユーザが巡航した Web ページの内容をそのまま保存することができる WebVCR というシステムを提案している<sup>11)</sup>。

これらの研究はいずれも新規性があり、興味深いものであるが、従来までの URL やタイトルに基づいたブックマークの概念を対象としている。本研究で提案するコンテキスト依存型ブックマークを利用することで、彼らが提案するシステムの利便性が向上するものと考えられる。

## 2.3 ユーザの振舞いに関する調査研究

Abrams らは、Web ブラウザやブックマーク管理システム設計のための基礎データを得ることを目的とし、ブックマークに対するユーザの価値観や位置付け等に関する分析を含めた、ブックマークに関する調査研究を行っている<sup>12),13)</sup>。

「ブックマークに対するユーザの意図」という着眼点は本研究と共通であるが、この調査研究結果に基づいた、新たなブックマークシステムの提案には至っていない。

### 3. コンテキスト依存型 Web ブックマーク

本章では、具体的なコンテキストデータの抽出方法について述べる。抽出するコンテキストデータは、1) ブックマーク行為の価値としてのブックマークランキング値、2) ブックマークの意味としての、ブックマークした Web ページとその他のページの特徴を表すキーワード（以下、特徴キーワードと略す）とする。

#### 3.1 ブックマークランキング値の算出(ブックマーク行為の価値の抽出)

検索エンジンで行われているランキング値の算出方法の多くは、Web ページの特徴やリンク構造等に基づいている。しかしながら、コンテキストブックマークのランキング値は単なる Web ページの価値ではなく、あくまでブックマーク行為の価値である。したがって、ランキング値はブックマークに至るまでのコンテキストデータを基に算出する。Web ページ自体の価値は、Web ページの内容やリンク構造が変化しない限り同一であるが、ブックマーク行為の価値は、ブックマークを付けたユーザや、そのときの状況によって大きく異なる。

本研究では、コンテキストブックマークのランキング値の算出を、以下の仮説に基づいて行う。

「ブックマークするまでのコンテキスト内において、ブックマークした Web ページとの類似度が高いページを多く閲覧していれば、そのブックマーク行為の価値は高い」。

すなわち、ブックマークまでに数多くの Web ページを閲覧していれば、ユーザは量的に詳しく調べたといえるので、ランキング値を高くする。また、ブックマークしたページとその他の閲覧したページの類似度が高ければ、ユーザの要求に近いページを多く閲覧しており、ユーザはシビアな選択を行ったといえるので、ランキング値を高くする。

図 3 に、この仮説のイメージ図を示す。

楕円形の領域を閲覧範囲とし、この閲覧範囲の中から選択したブックマークを旗に見立てている。円形内の矢印は、各 Web ページの特徴ベクトルを示しており、矢印の向きが近ければ、ページの特徴が近いことを示す。すなわち、閲覧範囲が広い（閲覧した Web ページ数が多い）方が、ランキング値は高く（図 3 上図）、ブックマークした Web ページに対して、閲覧した他のページの類似度が高い方が、ランキング値は高くなる（図 3 下図）ことを示している。

次に、ランキング値の算出式について説明する。ユーザが Web ページ集合  $S$  を閲覧した後に、ある Web

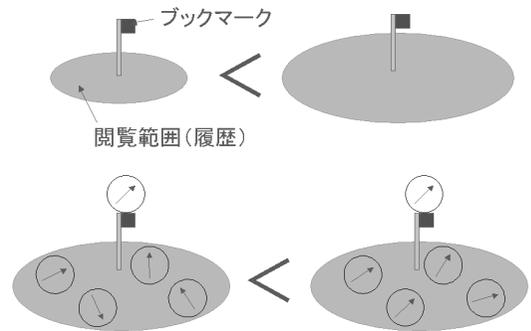


図 3 ランキング値算出のイメージ図

Fig. 3 Web bookmark and its context.

ページ  $x$  にブックマークを付けたという状況を想定する。ブックマークのランキング値を  $R_B$  とすると、 $R_B$  は以下の式で定義される。

$$R_B(x, S) = |S| \cdot \frac{\sum_{y \in S - \{x\}} \text{sim}(x, y)}{|S| - 1}$$

ここで、 $\text{sim}(x, y)$  は Web ページ  $x$  と  $y$  の類似度を表す。

この式の右辺は、“閲覧した Web ページ数”と“ブックマークした Web ページに対する閲覧した Web ページの類似度の平均値”の積である。すなわち、閲覧範囲が広く、ブックマークした Web ページに対して、閲覧した他のページの類似度が高ければ、ランキング値は高くなることを示している。

この式は、近似的に以下のようにも表すことができる。

$$R_B(x, S) = \sum_{y \in S - \{x\}} \text{sim}(x, y)$$

ランキング値算出の式は、この算出式が唯一というわけではない。しかし、計算の容易さから後述するプロトタイプでは、この算出式を採用している。

類似度  $\text{sim}(x, y)$  の算出方法としては、さまざまな手法が考えられるが、後述のプロトタイプシステムでは、各単語の出現頻度を基に各 Web ページの特徴ベクトルを作成して、コサイン相関値によって類似度を求める。

$$\text{sim}(x, y) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|}$$

ここで、 $\vec{x}$ 、 $\vec{y}$  は Web ページ  $x$ 、 $y$  の特徴ベクトルを表す。

なお、各 Web ページの特徴ベクトルの抽出方法は、まず、Web ページのテキストのソースファイルに対して、形態素解析を行い、名詞のみに対して、その出現頻度を算出する。そして、コンテキスト内のすべての

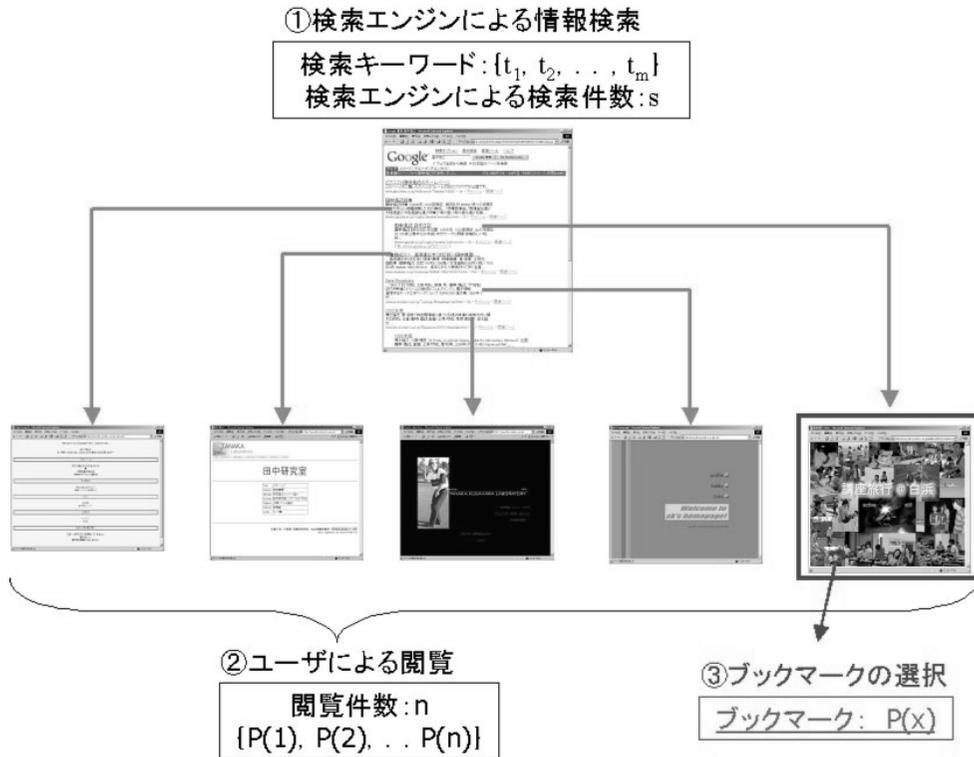


図4 ブックマークまでの流れ

Fig. 4 Process of context-dependent bookmarking.

名詞数を特徴ベクトルの要素数とし、各 Web ページにおける各名詞の出現頻度をその要素の値とする。このようにして、コンテキスト内の各 Web ページの特徴ベクトルを抽出する。

### 3.2 閲覧した Web ページの特徴抽出(ブックマークの意味の抽出)

Web 情報検索時のブックマークは、“コンテキストの中で気に入った Web ページ”と考えられる。したがって、“ブックマークしたページ”の特徴と、閲覧したが“ブックマークしなかったページ”の特徴の違いを明らかにすることにより、ユーザが“何が必要”で“何が不必要”であったかを明らかにすることができる。この特徴の違いこそが、ブックマークに対するユーザの判断基準であり、ブックマークの意味と考えることができる。

これを実現する方法としては、閲覧した各々の Web ページの特徴量を何らかの方法で抽出し、ブックマークした Web ページと、閲覧したその他の Web ページとの特徴的な違いを明確にするものである。

ここで、Web ページの特徴抽出のためのアルゴリズムについて説明する。検索エンジンに検索キーワードを与えて、得られる検索結果の中からいくつかの

Web ページを閲覧した後に、気に入ったページをブックマークした状況を想定する(図4参照)。ユーザが入力した初期検索キーワード群を  $\{t_1, t_2, \dots, t_m\}$  とする。また、検索結果の中から、ユーザが実際に閲覧した  $n$  個の Web ページ群を  $\{P(1), P(2), \dots, P(n)\}$  とする。この中でユーザが、ブックマークを付けた Web ページを  $P(x)$  ( $1 \leq x \leq n$ ) とする。

Web ページの特徴抽出の方法としては、TF・IDF 法<sup>14)</sup>を採用する。TF・IDF 法は、複数のドキュメントに対して、各々のドキュメントの特徴キーワードを抽出することに適した手法である。この TF・IDF 法においては、IDF の範囲をユーザが閲覧した Web ページ集合に限定することにより、各 Web ページ共通の特徴キーワードを排除することが可能である。したがって、各 Web ページ自体の特徴ではなく、閲覧範囲内の Web ページの特徴の違いを明確にすることができる。

ここで、Web ページ  $P(i)$  におけるキーワード  $t_j$  の TF・IDF 値を  $w_{ij}$  とすると、TF・IDF 値の算出式は以下ようになる。

$$w_{ij} = tf_{ij} \cdot \log \frac{n}{df_j}$$

Browsed Web pages	The Values of TF·IDF						
	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	W <sub>4</sub>	W <sub>5</sub>	W <sub>6</sub>	...
P(1)	0.2	1.3	0	0	0.4	0	...
P(2)	0.4	0	0	0.2	1.9	0	...
P(3)	0.2	1.2	0	0	0	0.2	...
P(4)	0.4	0	2.4	0	0	0	...
P(x)	0.2	0	0	2.8	0	2.1	...

図5 TF・IDFによる特徴抽出

Fig. 5 Extracting relative characteristics keywords using TFIDF.

ただし、

- $tf_{ij}$  は、Web ページ  $P(i)$  におけるキーワード  $t_j$  の出現回数、
- $df_j$  は、キーワード  $t_j$  が出現する Web ページ数、
- $n$  は、閲覧した Web ページ数、

である。

TF・IDF 法を用いて、各 Web ページの特徴抽出を行った場合の処理内容の例を図 5 に示す

P(1)~P(4) および P(x) は、閲覧した Web ページであり、W<sub>1</sub>~W<sub>6</sub> は、閲覧した Web ページ内のキーワードである。表中の数値は、各 Web ページにおける各キーワードの TF・IDF 値である。

ここで、各 Web ページ内で最も高い TF・IDF 値を持つキーワードを、その Web ページの特徴キーワードとする。すなわち、ブックマークした Web ページ、P(x) の特徴キーワードは“W<sub>4</sub>”とすることができる。

このようにして、ブックマークの際に、ユーザにとって、どのような特徴を持つ Web ページが必要で、どのような特徴を持つページが不要であったか」ということを表現することができる。

#### 4. プロトタイプの実装

3章のコンテキストデータの抽出方法に基づき、閲覧履歴を反映したコンテキストブックマークのプロトタイプを実装した。

プロトタイプシステムは、コンテキスト機能組み込み型のブラウザを開発することにより構築したが、実用段階では、セキュリティや利便性の問題から、プロキシサーバを利用する等といった、システムの構成についても検討が必要である。

システムが対象とするブックマーク時のユーザの行動モデルは、「検索エンジンに検索キーワードを与えて、その検索結果の中からいくつかの Web ページを閲覧した後に、気に入ったページにブックマークを付ける」という状況としている。その他のブックマーク

時の行動モデルについては、7章で述べる。

プロトタイプシステムの利用手順および想定している状況を以下に示す。

- ユーザはコンテキストブックマーク機能付きブラウザを利用し、検索エンジンに検索キーワードを与えて、検索する。
- 検索結果の中から、いくつかの Web ページを閲覧して、気に入ったものにブックマークを付ける。システムの動作手順を以下に示す。

- (1) ユーザが巡航しているサイトのうち、検索サイトのドメイン以外のものについては、そのページをユーザが閲覧していると見なし、そのページのソースコードを蓄積する。
- (2) ユーザがブックマークを行うと、ソースコードの収集を終了し、解析作業に入る。
- (3) 解析作業では、蓄積されたソースコードを解析し、ページごとに単語の出現頻度に基づいた特徴ベクトルを生成する。
- (4) ブックマークを付けたページの特徴ベクトルとその他のページの特徴ベクトルの類似度に基づいて、ブックマークランキング値を算出する。
- (5) TF・IDF 法により、各ページの特徴キーワードを抽出する。
- (6) ユーザに対して、検索キーワード、ブックマークランキング値、Web ページの特徴キーワードおよび URL を提示する。

すなわち、プロトタイプシステムの機能としては、コンテキストブックマークの取得までであり、コンテキストブックマークを利用するための機能は組み込まれていない。プロトタイプの外観は、通常のブラウザと大きな違いはない。なお、結果表示の際には、図 6 の最下部に表示されているウィンドウが現れる。

コンテキストブックマーク取得の例として、ワールドカップのチケットに関するホームページを探そうとした場合を考え、その流れ図を、図 6 に示す。

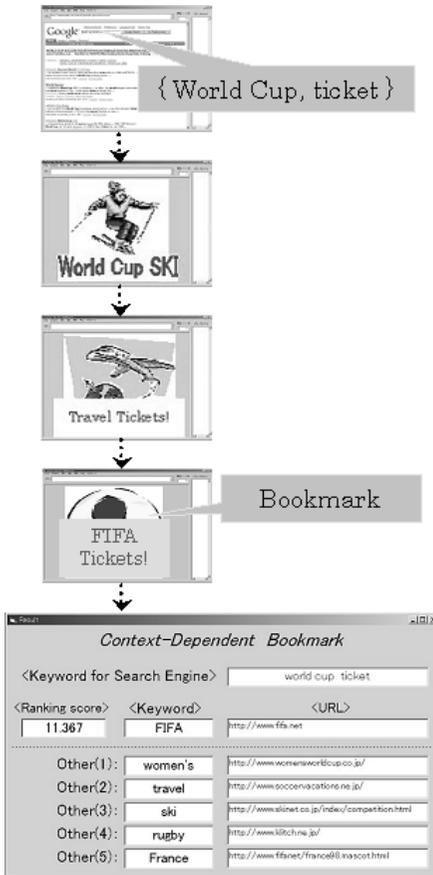


図6 コンテキスト依存型 Web ブックマーク (プロトタイプ)  
Fig. 6 Prototype of context-dependent bookmark.

ユーザが、たとえば {World Cup, Ticket} というキーワードで検索を行い、得られた検索結果の中から、いくつかのホームページを閲覧し、目的のページを見つけ、ブックマークする。するとシステムはブックマークした Web ページのランキング値と、閲覧した Web ページの特徴キーワードを提示する。

図6の例では、ユーザの目的および意図としては、サッカーのワールドカップのチケットに関する情報が欲しかったのだが、検索エンジンでは十分に絞り込むことができず、あまり関係のない“World Cup SKI”のページや、“Travel Tickets”のページ等にも迷い込んでしまったという例である。あわせて、図6では取得したコンテキストブックマークがメタデータとして {World Cup, Ticket, FIFA} のキーワードを保持し、そのランキング値が‘11.367’であるということを表している。また、この結果から、ユーザは {women’s, travel, ski, rugby, France} の特徴キーワードを持つ Web ページは不要であったことが分かる。

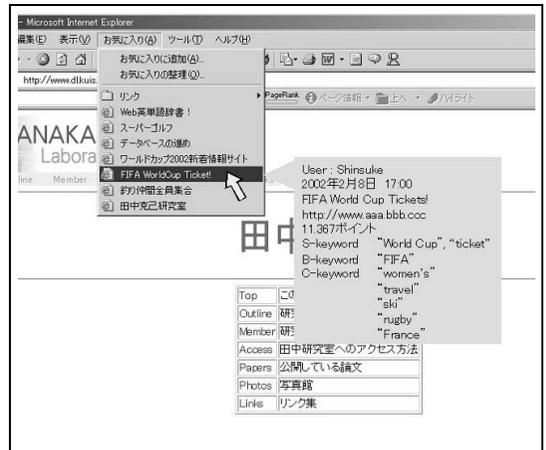


図7 Web ブックマーク機能の拡張のイメージ  
Fig. 7 Context-dependent bookmarks in Web browser.

### 5. コンテキストブックマークの用途

本章では、コンテキストブックマークの用途について述べる。

#### 5.1 ブラウザの Web ブックマーク機能

ブラウザの Web ブックマーク機能として、コンテキストブックマークを利用することにより、ユーザにブックマーク時の状況を容易に想起させることができる。図7にコンテキストブックマークをブラウザに組み込んだ場合のイメージを示す。

図7では、ユーザが保持しているコンテキストブックマークの1つに、マウスポインタを合わせると、このブックマークのコンテキストデータが、吹き出しによって表示されるという状況を示している。表示されるデータとしては、ユーザ名・ブックマークのタイトル・URL・ランキング値・日付・探索キーワード (S-Keywords) ブックマーク Web ページの特徴キーワード (B-Keywords) 非ブックマーク Web ページ (C-Keywords) の特徴キーワードがあげられる。

各コンテキストデータをユーザに対して表示することの効果をも、以下に示す。

- ランキング値  
ブックマーク取得時において、どの程度類似した Web ページを、どれだけ見て選んだのかの指針となる。
- 検索キーワード  
ブックマークを取得した際に閲覧した Web ページ群が、どのような検索キーワードによって検索された Web ページであったのかを確認することができる。これにより、意味的にどのような範囲

の Web ページをユーザが閲覧し、ブックマークしたのかということ容易に把握することができる。

- **ブックマーク Web ページの特徴キーワード**  
検索キーワードによって表現される、ブックマークした際に閲覧した Web ページ群全体の特徴に対して、ブックマークした Web ページは、どのような特徴を持っているのかを表現することができる。これによって、ユーザがどのような目的を持って Web ページを閲覧していたのかということ、直感的に把握することができる。
- **非ブックマーク Web ページの特徴キーワード**  
検索キーワードによって表現される、ブックマークした際に閲覧した Web ページ群全体の特徴に対して、ブックマークされなかった各々の Web ページは、どのような特徴を持っているのかを表現することができる。すなわち、どのようなキーワードに代表される Web ページが、ユーザにとって不要であったのかということ、直感的に把握することができる。

実装したプロトタイプシステムでは、図 7 に示したコンテキストブックマークの利用機能のブラウザへの組込みはまだ行っていない。

この Web ブックマーク機能としてのコンテキストブックマークの用途は、単にブックマーク時の状況の想起を容易にすることのみではない。ブックマークがコンテキストデータというメタデータを保持しているため、このデータを利用した整理および表示が可能となる。たとえば、以下のような機能が考えられる。

- ランキング値・特徴キーワード・日付等によるソーティング
- 各特徴キーワードに基づいたブックマーク間の関連の視覚化
- コンテキストデータに基づいたブックマーク検索  
これらの機能は特にブックマークが大量になったときに有効である。

## 5.2 コンテキストブックマークの問合せとしての利用

本節では、コンテキストブックマークの問合せとしての利用について述べる。コンテキスト依存型ブックマークは、コンテキストデータとして、以下のキーワードを保持している。

- 検索エンジンに与えた検索キーワード
  - 非ブックマーク Web ページの特徴キーワード
  - ブックマーク Web ページの特徴キーワード
- これらのキーワードから「どのような意味的領域の Web ページを閲覧し」、「どのような特徴キーワードを

持つ Web ページを閲覧して」、「最終的にどのような特徴キーワードを持つ Web ページを選択したのか」ということを表現することが可能である。

つまり、ブックマーク取得時のユーザの判断基準を抽出できるため、このコンテキストデータを利用することにより、ブックマークした URL にとらわれず、ユーザの意図を反映した Web ページを探索するための問合せを生成することができる。

具体的な問合せ方法は、ブックマークした Web ページの特徴ベクトルを基準とし、コンテキスト範囲内の Web ページの特徴ベクトルとの差異を増幅することにより、ユーザの要求に近いと思われる特徴ベクトルを生成する。さらに、問合せを適用する未閲覧領域の各 Web ページの特徴ベクトルと、コンテキストブックマークから生成した特徴ベクトルの類似度を計算し、類似度の上位のものをピックアップするという方法である。

ここでコンテキストデータに基づく特徴ベクトルの差異増幅の方法について説明する。

まず、ブックマークした Web ページの特徴キーワード（ユーザの要求に合致した Web ページの特徴キーワード）に対しては、特徴ベクトルの中の該当する要素の重みを大きくする。重みの増幅量については、ブックマークした Web ページ中のこの特徴キーワードの出現頻度と他の Web ページ中での出現頻度の差異に基づいて決定する。

また、コンテキスト範囲内の各 Web ページの特徴キーワード（すなわちユーザの要求には合致しない Web ページの特徴キーワード）に対しては、特徴ベクトルの中の該当する要素の重みを小さくする。重みの減少量については、各 Web ページにおける、その特徴キーワードの出現頻度とブックマークした Web ページにおける出現頻度の差異に基づいて決定する。

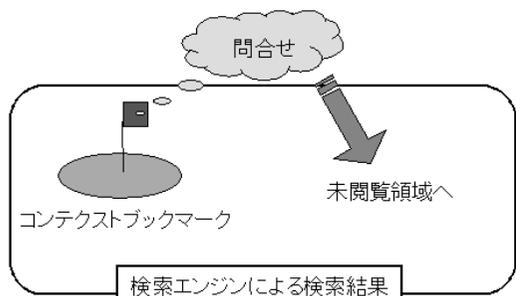
このコンテキストブックマークの問合せ利用の有用性の評価については、次章で述べる。

以下に、ブックマークの問合せとしての具体的な利用方法について述べる。

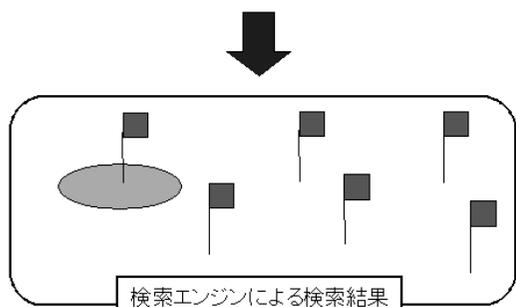
### ● 未閲覧領域の検索

#### (1) 検索結果中の未探索領域への適用

検索エンジンを利用して、ユーザが必要な情報を検索する場合に、検索結果として数万件もしくはそれ以上の Web ページがヒットすることがある。この場合、ユーザがすべての Web ページをチェックすることは不可能に近く、閲覧した Web ページの範囲の中から気に入った Web ページを



1. コンテキストブックマークからの問合せ生成
2. 未閲覧領域への問合せの適用



3. ユーザの要求に合致した Web ページのピックアップ

図 8 コンテキストブックマークによる問合せの例

Fig. 8 Instance of utilizing context-dependent bookmarks.

採用することになる。すなわち、未閲覧の Web ページの中にユーザにとって価値の高いページが存在したとしても、その Web ページを利用することはできない。このような状況に対して、コンテキストブックマークによる問合せをこの検索結果中の未探索領域に適用することにより、数万件の未閲覧 Web ページの中からユーザの要求に合致した Web ページをピックアップすることが可能である（図 8 参照）。

## (2) 他の探索領域への適用

ユーザが取得したコンテキストブックマークに基づく問合せを、意図的に他の領域（コンテキストブックマークのコンテキスト範囲ではない領域）へ適用することが可能である。いくつかの例を以下に示す。

- (a) ほぼ同様な特徴の Web ページを数多く収集したい場合。たとえば、検索エンジンを利用して情報検索を行い、ユーザの目的に合致した Web ページをいくつかブックマークした

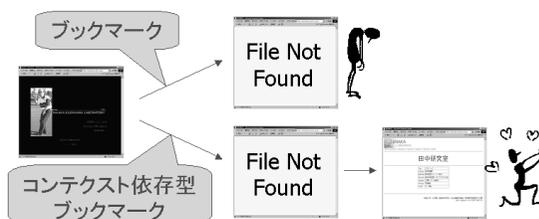


図 9 期限切れ URL の代替ページの問合せ

Fig. 9 Context-dependent bookmark as query.

が、同じような特徴の Web ページをさらに別の検索エンジンで検索したいという状況を考える。このような場合、はじめに取得した Web ページをコンテキストブックマークとして残し、別の検索エンジンの検索結果に対して適用することにより、同様な情報を数多く取得することができる。

- (b) ある条件を満たす情報に対して、この条件の一部を変更した情報を検索したい場合。たとえば、{九州, 温泉地} という検索キーワードで検索した検索結果の中から、ユーザの好みにあった温泉地に関する Web ページをコンテキストブックマークとして残す。ユーザは、このコンテキストブックマークから生成される問合せを、{四国, 温泉地} というキーワードで検索した検索結果に対して適用することにより、ユーザの好みにあった四国の温泉地に関する情報を取得することが可能となる。

## ● URL が有効期限切れとなった場合の代替ページの検索

URL が変更されてアドレスが不明になった場合、従来のブックマークはその価値を失うが、本ブックマークの場合には、以前にブックマークした Web ページに近い内容の Web ページを問い合わせることが可能である（図 9 参照）。

## 6. コンテキストブックマークの有用性評価

5.2 節において、コンテキストブックマークの問合せ利用について述べた。本章ではこのブックマークの問合せ利用に関する評価実験を行うことにより、コンテキストブックマークの有用性の評価を行う。

未閲覧領域内の Web ページ群の中から、ユーザの

要求に近い Web ページをピックアップする方法としては、コンテキストデータを考慮せずに、単にブックマークしたページの特徴ベクトルを利用する方法がある。つまり、このブックマークしたページの特徴ベクトルと、未閲覧領域の Web ページとの類似度を算出し、類似度が上位のページをピックアップする方法である。

この方法に対して、コンテキストブックマークの問合せ利用では、ブックマークした Web ページの特徴ベクトルを基準とし、ユーザの意図が反映されているコンテキストデータに基づいた差異増幅が行われている。したがって、未閲覧領域の Web ページに対して、ブックマークした Web ページの特徴ベクトルによる問合せに比べると、コンテキストデータに基づいた差異増幅特徴ベクトルによる問合せは、ユーザの意図をより反映した Web ページをピックアップすることができるはずである。逆にいえば、ユーザの意図をより反映した Web ページをピックアップできるということを証明できれば、コンテキストブックマークによりユーザの意図を抽出することが可能であるといえる。評価実験の手順を示す。

- (1) ユーザが、検索エンジンに検索キーワード  $K$  を入力して、情報検索を行う。
- (2) 検索結果の中からいくつかの Web ページを閲覧し、気に入った Web ページ  $X$  をコンテキストブックマーク  $B_X$  として保存する。
- (3) 検索結果の中で、このとき閲覧しなかった Web ページの中から  $N$  件の未閲覧 Web ページを取り出す。そして、ユーザがブックマークしたときと同じ判断基準でこの  $N$  件の Web ページを判定して、ユーザの目的と合致するページ ( $M$  件) にチェックを付ける。
- (4) ブックマークした Web ページ  $X$  の特徴ベクトルと、コンテキストブックマーク  $B_X$  に基づいた差異増幅特徴ベクトルを生成し、未閲覧領域内の  $N$  件の Web ページとの類似度を算出する。
- (5)  $N$  件の Web ページを類似度が高い順序で並べ替え、上位  $R$  件の Web ページをユーザに対して提示する。
- (6) システムからユーザに提示される  $R$  件の中に含まれる、ユーザの目的に合致した (ユーザがチェックを付けた) Web ページの数によって、適合率と再現率を算出する。

各問合せ方式の評価基準として、適合率と再現率を採用した。本評価実験における適合率と再現率の定義

を以下に示す。ただし、ユーザが目的と合致していると判断してチェックを付けた Web ページの数を  $M$  件、システムが各問合せ方式においてユーザに提示する Web ページの件数を  $R$  件としている。

適合率 = 各問合せ方式において、ユーザに提示される  $R$  件の Web ページのうち、ユーザの目的に合致した Web ページの数の割合

再現率 = ユーザの目的に合致した  $M$  件の Web ページのうち、ユーザに提示される  $R$  件の Web ページに含まれる Web ページの割合

本評価実験では、被験者が上述の手順に基づいて、さまざまな話題を含むキーワードにより Web 検索を行い、その中からある特定の話題について書かれた Web ページをコンテキストブックマークとして取得することとした。

(例)

- { 地球温暖化, 二酸化炭素 } というキーワードで検索し、地球温暖化のメカニズムについて書かれたページ
- { 狂牛病 } というキーワードで検索し、政府の政策について書かれたページ

等、8 種類のコンテキストブックマーク。

以上の方法で取得したコンテキストブックマークから、ブックマークした Web ページの特徴ベクトルと、コンテキストデータに基づいた差異増幅特徴ベクトルを生成した。そして、ブックマーク取得時に閲覧した Web ページと同じ検索キーワードにより検索された  $N$  件の未閲覧 Web ページ (本実験では  $N=20$ ) と、これら 2 つの特徴ベクトルとの類似度を算出して、ランク付けを行った。この結果に対して、ユーザに提示する件数 ( $R$ ) を 1~10 まで変化させて、この件数に応じた適合率と再現率を、各問合せ方式により比較した。比較した適合率と再現率の値は 8 種類のデータの平均値を用いた。適合率に関する比較評価結果を図 10 に、再現率に関する比較評価結果を図 11 に示す。

ただし、

オリジナル：ブックマークした Web ページの特徴ベクトルとの類似度によるランク付け (コンテキストブックマークを利用せず)  
 差異増幅：コンテキストデータに基づいた差異増幅特徴ベクトルとの類似度によるランク付け (コンテキストブックマークを利用) である。

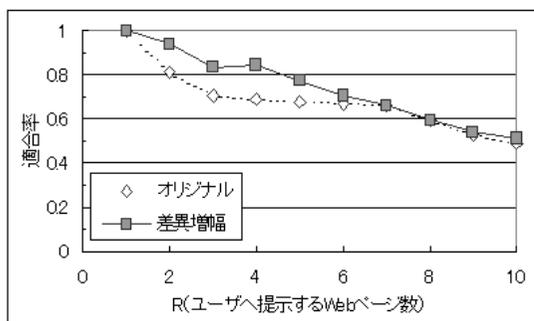


図 10 比較評価結果 (適合率)

Fig. 10 Result of comparison on relevance factor.

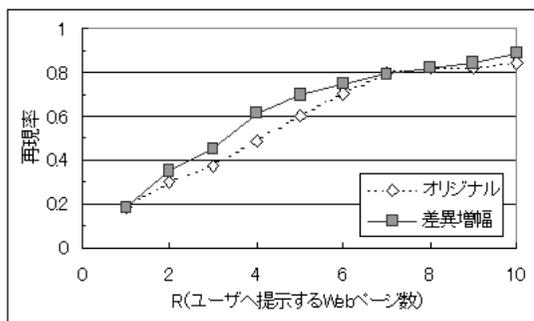


図 11 比較評価結果 (再現率)

Fig. 11 Result of comparison on recall factor.

適合率に関しては、Rの値が大きくなるにつれ、オリジナル、差異増幅ともに減少した。また、差異増幅は、つねにオリジナル以上の適合率を示した。また、図 10 において、Rの値が大きくなるとオリジナルと差異増幅の差がなくなっている。これは、ユーザの目的に合致した Web ページの数よりも、Rの値（ユーザに提示する Web ページ数）が大きくなると、ユーザの目的に合致した Web ページの順位が多少悪くても、R 件に含まれてしまうからである。

再現率に関しては、Rの値が大きくなるにつれ、オリジナル、差異増幅ともに増加した。また、差異増幅は、つねにオリジナル以上の再現率を示した。また、図 11 において、Rの値が大きくなるとオリジナルと差異増幅の差がなくなっている。これは適合率の場合と同様で、ユーザの目的に合致した Web ページの数よりも、Rの値（ユーザに提示する Web ページ数）が大きくなると、ユーザの目的に合致した Web ページの順位が多少悪くても、R 件に含まれてしまうからである。

以上の結果から、単にブックマークした Web ページの特徴ベクトルよりも、コンテキストデータに基づいた差異増幅特徴ベクトルの方が、よりユーザの意図

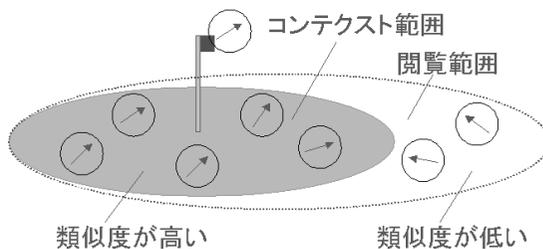


図 12 無関係な Web ページの排除

Fig. 12 Scope of the context.

を反映することができるという。

コンテキストブックマークの評価実験としては、増幅率の最適化を含めた差異増幅の方法等、さらに詳しく検討すべき要件があるが、本実験によりコンテキストブックマークの有用性を示すことはできたといえる。

## 7. ユーザの行動モデルを考慮したコンテキスト範囲の限定方法

ここまでは「検索エンジンに検索キーワードを与えて、その検索結果の中からいくつかの Web ページを閲覧した後に、気に入ったページにブックマークを付ける」という状況をコンテキストブックマーク取得時の前提としてきた。しかしながら、実際のブックマーク時のユーザの行動モデルとしては、いろいろなパターンが考えられる。それぞれの状況におけるコンテキスト依存型ブックマークの適用について検討する。

- 自由にナビゲーションした後に、ブックマークする場合 (閲覧範囲内の無関係なページの扱い) 検討すべき問題としては、①ナビゲーション自体に明確な目的があるとは限らず、最終的にブックマークしたページとは無関係なページを閲覧する可能性がある、②リンクをたどりながらナビゲーションする場合、パスの関係上、無関係なページを閲覧する可能性がある、等があげられる。したがって、このような場合、ブックマークした Web ページとは無関係な Web ページをコンテキストから排除する。具体的には、ブックマークした Web ページと無関係な Web ページは、ブックマークした Web ページとの類似度が低いと考えられるため、ある閾値以下の類似度の Web ページについては、コンテキスト範囲から排除するという方法を適用する (図 12 参照)。
- 自由にナビゲーションした後に、ブックマークする場合 (同ドメイン内の Web ページの扱い)

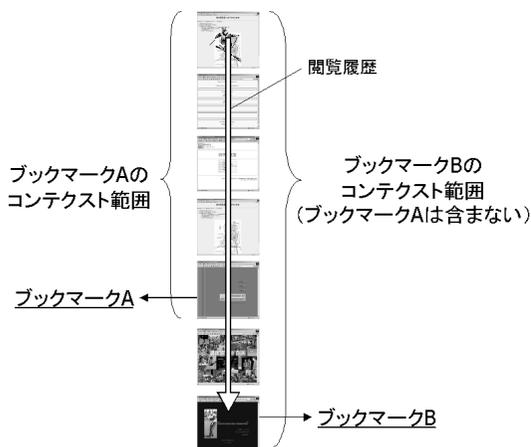


図 13 コンテキスト範囲の重複  
Fig. 13 Overlap of the context.

ブックマークした Web ページと同ドメインの Web ページが、コンテキスト範囲に含まれる場合を考える。この Web ページが、ブックマークした Web ページよりも、ドメイン内のトップページに近ければ（階層が上位であれば）、この Web ページはユーザにとって“不必要なページ”ではなく、目的のページにたどり着くための通過点である可能性が高い。したがって、ブックマーク以外の同ドメイン内の Web ページは、コンテキスト範囲から排除する。

#### ● 複数の Web ページをブックマークする場合

1 つのセッションの中で、複数の Web ページをブックマークした場合、コンテキスト範囲の限定が問題となる。検討すべき問題としては、①ブックマークした Web ページは他のブックマークのコンテキストに含まれるか、②ある Web ページをブックマークした後に、連続してナビゲーションを行いブックマークする場合、2 つ目のブックマークのコンテキスト範囲をどのように設定するのか、等があげられる。

まず、①について、ブックマークされた Web ページは“ユーザにとって不要”ではないため、他のブックマークのコンテキスト範囲から排除する。②については、ブックマークごとに途切れるわけではない。したがって、2 つ目のブックマークのコンテキスト範囲は、1 つ目のブックマークのコンテキスト範囲も含むが、ブックマークされる Web ページと無関係な Web ページはコンテキスト範囲から排除する（図 13 参照）。

#### ● コンテキストと無関係な Web ページをブック

#### マークする場合

まったく別の目的でナビゲーションを行っている場合に、たまたま発見した Web ページをコンテキストとは無関係にブックマークする場合について検討する。ただし、ブックマークがコンテキストとは無関係ということシステムが自動認識するための手段がない。したがって、このような場合にはユーザがブックマークに対して注釈を付けることにより、システムに認識させることにする。

#### ● コンテキスト範囲に“リンク集である Web ページ”を含む場合

リンク集は、他のページへナビゲーションするためのページという意味合いが強いページであり、コンテキスト範囲に含めるべきではない。したがって、リンク集であると認識されるページはコンテキスト範囲から排除する。

ここで、Web ページがリンク集であるかどうかの判定は、ページ内の文字数に対するリンク数の割合が、ある閾値を超えるか超えないかにより行うものとする。

#### ● “リンク集である Web ページ”をブックマークした場合

リンク集は、他のページへナビゲーションするためのページという意味合いが強いページであり、リンク集ではない Web ページとは区別する必要がある。また、リンク集であることでコンテキストブックマークとして有効利用できる可能性もある。したがって、通常コンテキスト依存型ブックマークと区別するために、コンテキストデータの中にリンク集であるというメタデータを追加する。

## 8. おわりに

本研究において得られた成果と知見を以下に示す。

- Web 利用時の成果と考えられるブックマークに至るコンテキスト（閲覧履歴）を取得し利用する、“閲覧履歴を反映したコンテキスト依存型 Web ブックマーク”の概念を提案した。
- この“閲覧履歴を反映したコンテキスト依存型 Web ブックマーク”に関するアルゴリズムを設計し、プロトタイプの実装を行った。Web ブックマークに付加するコンテキストデータとしては、1) コンテキストに基づくランキング値、2) ブックマークした Web ページとその他のページとの違いを表す特徴量、を提案した。
- 本システムの利用用途および効果として、以下の

内容について検討した。

- ブラウザのブックマーク機能の向上
- ブックマークの間合せとしての利用
- コンテキストブックマークによる間合せ方式に関する評価試験を行い、コンテキストブックマークの有用性を確認した。

また、コンテキストブックマークの検討項目を以下に示す。

- (1) Web ページの特徴付け手法の検討  
ブックマークした Web ページと、閲覧した他の Web ページとの特徴の違いを明らかにするために、各 Web ページの特徴付けを TF・IDF 法によって行っている。本手法では 1 度に処理する文書数が少ないために、IDF 値にノイズが乗りやすいという性質を持つといえる。また、TF・IDF 法はテキスト文に対する特徴付けに用いる手法であるために、Web ページの特徴付けに用いる手法として十分であるとはいえない。したがって、本手法で扱う、少数かつマルチメディアデータである Web ページの特徴付けの手法として、どのような手法が適切なのかということについても、検討すべきと考える。
- (2) 特徴キーワード抽出方法に関する検討  
プロトタイプによりコンテキストブックマークを取得する実験を行った中で、理解しがたい特徴キーワードが抽出されることも多かった。理由としては、上述のように TF・IDF 法による各 Web ページの特徴付けが必ずしも適切とはいえないことも関係していると考えているが、そもそも特徴キーワードとは認めがたい“年”、“月”、“日”、“円”等が抽出されることも多いので、不要語の設定や形態素解析手法の改善についても検討すべきと考える。
- (3) 閲覧順序の解析に関する検討  
現在のプロトタイプでは、コンテキストデータの抽出において、ユーザによる Web ページの閲覧順序は考慮していない。検索エンジンの検索結果からのみ閲覧する場合は、閲覧順序にはほとんど意味がないと考えられるが、自由にナビゲーションする場合には、ブックマークに関する何らかの意味を含んでいる可能性もある。したがって、ユーザによる閲覧順序の扱いについては、今後の検討課題とする。
- (4) ユーザの満足度取得の検討  
現在のプロトタイプでは、ブックマークした Web ページに対するユーザの満足度を反映さ

せることができない。ブックマークに対するユーザの満足度は、そのコンテキストだけでなくユーザによって異なるものである。したがって、ユーザを判別するデータ (ID もしくはプロフィール) とともに、ユーザの入力により得られるユーザの満足度もブックマークのメタデータとして付加することも検討する。

今後は、さらなるシステムの検証実験を行うとともに、機能の向上について検討する予定である。

謝辞 本研究の一部は、文部科学省科学研究費特定領域研究 (C) (2) 「Web の意味構造発見に基づく新しい Web 検索サービス方式に関する研究」(課題番号: 14019048)、および基盤研究 (A) (2) 「モバイル環境におけるコンテンツのマルチモーダル検索・呈示と放送コンテンツ生成」(課題番号: 14208036) による。ここに記して謝意を表します。

#### 参 考 文 献

- 1) 中島伸介, 黒田慎介, 田中克己: Web 情報検索における文脈依存型ブックマーク, 情報処理学会第 63 回全国大会 (2001).
- 2) 中島伸介, 黒田慎介, 田中克己: 閲覧履歴を反映したコンテキスト依存型 Web ブックマーク, Proc. DBWeb2001, 情報処理学会シンポジウムシリーズ, Vol.2001, No.17, pp.201-208 (2001).
- 3) Maarek, Y.S. and Ben-Shaul, I.: Automatically Organizing Bookmarks per Contents, *5th International World Wide Web Conference* (1996).
- 4) Schmiedel, A. and Volle, P.: Using Structured Topics for Managing Generalized Bookmarks, *5th International WWW Conference* (1996).
- 5) Keller, R.M., Wolfe, S.R., Chen, J.R., Rabinowitz, J.L. and Mathe, N.: A Bookmarking Service for Organizing and Sharing URLs, *Proc. 6th International WWW Conference* (1997).
- 6) Takano, H. and Winograd, T.: Dynamic Bookmarks for the WWW, *Proc. 9th ACM Conference on Hypertext and Hypermedia* (1998).
- 7) Li, W., Vu, Q., Agrawal, D., Hara, Y. and Takano, H.: PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management, *Proc. 8th International WWW Conference* (1999).
- 8) Ljungstrand, P., Redstrom, J. and Holmquist, L.E.: WebStickers: Using Physical Tokens to Access, Manage and Share Bookmarks to the Web, *Designing Augmented Reality Environments (DARE'2000)* (2000).
- 9) Hascoet, M.: A User Interface Combining

- Navigation Aids, *Proc. 11th ACM on Hypertext and hypermedia*, pp.224–225 (2000).
- 10) Kaasten, S. and Greenberg, S.: Integrating Back, History and Bookmarks in Web Browsers, *Extended Abstracts of the ACM Conference of Human Factors in Computing Systems (CHI'01)* (2000).
- 11) Anupam, V., Freire, J., Kumar, B. and Lieuwen, D.F.: Automating Web navigation with the WebVCR, *Proc. 9th International WWW Conference* (2000).
- 12) Abrams, D., Baecker, R. and Chignell, M.: Information Archiving with Bookmarks: Personal Web Space Construction and Organization, *Conference on Human Factors and Computing Systems*, pp.41–48 (1998).
- 13) Abrams, D. and Baecker, R.: How People Use WWW Bookmarks, *CHI97 Electronic Publications: Late-Breaking/Short Talks* (1997).
- 14) Salton, G. and Yang, C.S.: On the specification of term values in automatic indexing, *J. Documentation*, Vol.29, No.4, pp.351–372 (1973).

(平成 13 年 12 月 20 日受付)

(平成 14 年 3 月 27 日採録)

(担当編集委員 石川 博)



中島 伸介(学生会員)

1995 年神戸大学工学部生産機械工学科卒業。1997 年神戸大学大学院自然科学研究科機械工学専攻博士前期課程修了。2000 年京都大学大学院工学研究科環境工学専攻受託研究員。2001 年京都大学大学院情報学研究科社会情報学専攻博士後期課程入学、現在に至る。



黒田 慎介

2000 年京都大学工学部電気工学科卒業。2002 年京都大学大学院情報学研究科社会情報学専攻修士課程修了。2002 年(株)エヌ・ティ・ティ・ドコモ入社、現在に至る。



田中 克己(正会員)

1974 年京都大学工学部情報工学科卒業。1976 年京都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 年同大学工学部教授(情報知能工学専攻)。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、IEEE Computer Society、ACM 等各会員。