

Title	利用者特性とドキュメント特性を利用したブラウジング環境
Author(s)	金井, 秀明; 箱崎, 勝也; 石川, 克則
Citation	情報処理学会論文誌 : データベース, 41(SIG3): 46-57
Issue Date	2000-05-15
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/4558
Rights	<p>社団法人 情報処理学会, 金井秀明, 箱崎勝也, 石川克則, 情報処理学会論文誌 : データベース, 41(SIG3), 2000, 46-57. ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.</p>
Description	

利用者特性とドキュメント特性を利用したブラウジング環境

金井 秀明[†] 箱崎 勝也[†] 石川 克則^{††}

従来、電子図書館における検索として、書誌情報などのキーワードによる検索が利用されている。しかし検索対象が明確でなく漠然と興味のある情報を探す場合、実際に図書館内を散策しながら図書を探す必要がある。このような方法はブラウジング検索と呼ばれ、電子図書館への適用が試みられている。効率的に散策を行うには、利用者の好みを反映した情報オブジェクトを配置する必要がある。しかし厳格に好みを反映した場合、予め興味のあることが明確な情報以外はフィルタリングされ、興味のありそうな情報を偶然見つけることを阻害する可能性がある。

本研究では、散策の初期段階から情報オブジェクトをフィルタリングよりも、情報オブジェクトが示しているコンテンツの特性と利用者特性の関係を利用者に提示することで、散策支援を目指す。すなわち、利用者の参照履歴、興味特性や利用者とコンテンツの類似度特性の視点から散策空間を提示する。利用者がその空間を眺めることで、空間内の情報オブジェクトがどのような特性を持っているかを把握できる。そして散策過程の特性を利用者特性にフィードバックする。その変更した利用者特性に基づいて散策空間を更新する。このようにして利用者特性に適合した散策空間を提供する。本論文ではこの散策手法を提案し、その試作システムについて述べる。

A Browsing Environment using Features of Readers and Document

HIDEAKI KANAI,[†] KATSUYA HAKOZAKI[†] and KATSUNORI ISHIKAWA^{††}

The current digital libraries usually adopt the searching by bibliographies and some keywords. When users do not decide specific target books and check whether some interesting books exist at the libraries or not, they must stroll in real libraries. Now, the strolling is being applied to information retrieval in digital libraries, and the method is called browsing. In order to practice effective strolling, it is necessary to restructure information space according to user's interests. However, if the user's interests is rigorously reflected in the structure, it would prevent users from finding an interesting-like information by chance.

We focus on the method of making users aware of relation between user features and information features rather than the strict information filtering. In this paper, we presents a browsing environment using features of readers and documents. In this environment, the users becomes aware that there is information which they seem to be interested in, from objects in information space, and stroll there. The objects visualize user feature, information feature, user's strolling action history. Moreover, the configuration of objects is adaptable to change in each user's feature.

1. はじめに

近年のインターネットの急速な普及により、WWW (World Wide Web) 情報資源から利用者の目的にかなう情報を検索することが重要になっている。従来よりWWW 情報や電子図書館に代表される情報検索では、カテゴリー分類やキーワードによって検索した結果か

ら、利用者が目的にかなうものを選択することが行われてきた。利用者が自身で明確に目的とする情報を把握している場合、この方法によって効率的に必要な情報を検索することが可能である。しかし、検索対象が明確でなく漠然と興味のある情報を探したい場合には必ずしも適していない。そのため、例えば図書検索の場合、実際に図書館内を散策し興味のありそうな図書を手にとって眺めながら図書を探す必要がある。この方法はブラウジング検索と呼ばれ、ショッピングや Web ページの閲覧で行われている。また、現在ブラウジング検索を電子図書館に適用する試みが行われている。

一方、散策空間の規模が大きくなるにつれ、その空

[†] 電気通信大学大学院情報システム学研究所
Graduate School of Information Systems, University of
Electro-Communications

^{††} 凸版印刷株式会社 マルチメディア事業部
Multimedia division TOPPAN Printing Co.,Ltd

間を全て散策することは難しく、結果的に有用な情報を探ることが困難になる。効率的に散策を行うために、予め利用者の好みを反映した空間を生成する必要がある。しかし厳密に好みを反映した場合、偶然興味のありそうな情報を探ることを阻害することになりかねない。

そこで本研究では、散策の初期段階から利用者の好みに基づいた厳密なフィルタリングを行うのではなく、むしろ散策空間にある情報がどのような特性を持ったものであるのかを利用者に提示することで散策支援を試みる。以下、本論文では散策対象の情報のことを総称して「ドキュメント」と呼ぶ。ただしドキュメントは図書に限定したものではない。

筆者らは以前、個人利用に適合した仮想図書館「やわら」の構想を提案した¹⁾。本論文では、特に、利用者特性とドキュメント特性の関係を利用者が容易に把握できるブラウジング手法を提案し、試作システムについて述べる。以下、2章では従来研究と比較し、提案手法の概要について述べる。3章、4章で提案手法について説明する。5章では試作システムの構成を述べ、6章ではその実行例を紹介しシステムの有効性を示す。

2. 利用者情報とドキュメント特性によるブラウジング手法

2.1 散策支援

散策空間で利用者がドキュメント特性を容易に把握するには、利用者の興味のありそうなドキュメントを推定し、その推定結果を反映した散策空間を提示することが重要である。そこで本研究では以下の2点について検討を行う。

- 利用者特性の表現
- 散策空間の構成

2.1.1 利用者特性の表現

ブラウジング検索における利用者特性としては、「静的要素」と「動的要素」の2要素が考えられる。

- 静的要素: 予め利用者が登録した興味分野や事項²⁾。
- 動的要素: 散策過程で現れるもの。例えばドキュメントの参照/操作履歴や、参照したドキュメントに対する利用者の興味度。

ブラウジング検索では静的要素に合致したドキュメントだけでなく、散策を通して見つけたドキュメントも有用なものである。また、予め利用者が自身の読書傾向を正確に把握しているとは限らず、さらに散策を通して読書傾向に変化が生じる場合も考えられる。そこで動的要素を静的要素に反映させる形で利用者特性を扱い、散策過程での特性を考慮して利用者にドキュ

メントを提示する手法が研究されている。例えばインターネット新聞システムでの参照/操作履歴情報を反映させたもの³⁾や利用者の入力した興味度情報を反映させたもの⁴⁾があり、その有効性が報告されている。しかし、これらのシステムでは参照/操作履歴情報と興味度情報を統合的に扱っていない。そのため参照履歴情報だけを用いた場合、興味はなくともだた参照頻度が高いことだけで利用者が興味持っていると思われる。また興味度情報だけを用いた場合、利用者の興味度入力が必要となり、利用者の負担が大きくなるという問題点がある。

2.1.2 散策空間の構成

従来より散策空間の構成に関する研究では、ドキュメントをサムネイルアイコンとして表現し、それを3次元空間上に配置することが行われている。例えばWebページを仮想的な本に整形し、それを3次元仮想デスクトップ上に配置する3次元Webブラウザシステムがある⁵⁾。また利用者にその仮想的な本を配置させ、ドキュメント管理を行うシステムがある⁶⁾。これらのシステムでは利用者がドキュメントを操作したり、ドキュメントを検索することができる。しかしこれらのシステムでは、散策空間やドキュメントを表現したもの(以下、オブジェクトと呼ぶ)の構成にドキュメント同士の類似性や利用者特性を利用していない。そのため、利用者がドキュメントの特性を把握するのが難しいという問題点がある。

一方、それらの特性情報を散策空間の構成に利用したシステムがいくつかある。例えば、多変量解析手法の一つである数量化3類⁷⁾によって求めたドキュメント同士の類似性を使ってオブジェクトを配置するシステムがある⁸⁾。また利用者の興味とドキュメントとの類似性を使ってオブジェクトを配置するシステムがある⁹⁾。しかし、これらのシステムでは予め入力した利用者特性(静的要素)に基づいて散策空間を構成するため、散策過程を考慮した利用者特性を導入しておらず、散策過程を通じた読書傾向の追従性に関して問題がある。

2.2 提案する手法

上記の点に関して、本研究では以下のように扱う。また本論文では利用者特性の静的要素と動的要素を各々「読書特性」、「散策特性」と呼ぶ。

- 散策特性として「参照履歴」と「参照したドキュメントに対する興味度」の両者を統合的に扱う。それらを予め登録された読書特性にフィードバックすることで、散策過程を考慮した利用者特性表現を導入する。これにより、一方だけの特性情報に

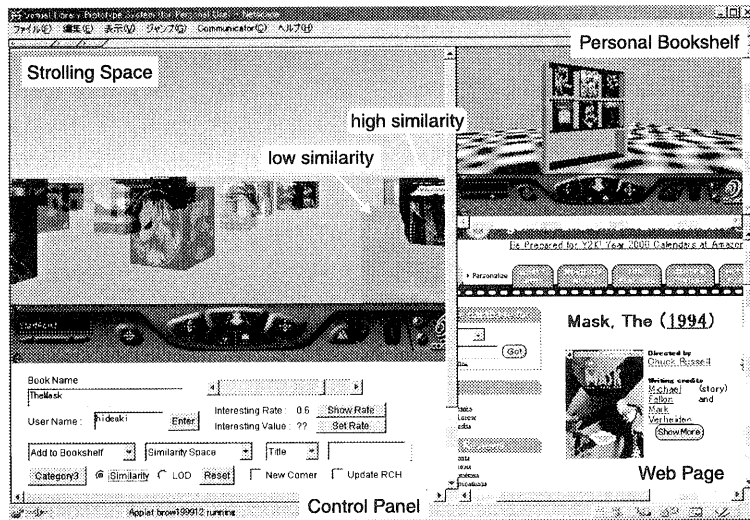


図 1 試作システムの実行画面

Fig. 1 Display image of a prototype system.

依存することなく、利用者特性を表現することが可能となる。

- 利用者に上記の利用者特性やドキュメント特性の観点から散策空間を提示する。利用者がその空間を眺めることで、空間内の情報がどのような特性を持っているかを把握できるようにする。そして散策特性を読書特性にフィードバックすることで読書特性を更新し、それに基づいて散策空間を更新する。このようにして、より利用者特性に適合した散策空間を提供する。

上記のブラウジング環境を提案し、図 1 に示すような試作システムを開発した。以降、本手法の詳細について述べる。

3. ドキュメント特性と利用者特性

本研究でのドキュメント特性、読書特性および散策特性について述べる。

3.1 ドキュメント特性

ドキュメントはその書誌的事項(著者、題名、雑誌名、メディアタイプ)とドキュメント特性によって表現される。ドキュメント特性の表現には、従来よりドキュメントを特徴付けるキーワードやカテゴリを属性とし、その属性に対する重みによるベクトルモデルが用いられている。本研究でも同様の方法を用いる。すなわち、式 1 に示すようなドキュメント D_i に対するドキュメント特性ベクトルを用いる。キーワードの選定および重みの割り当て方法として、本研究では人手による方法を採用する。詳細は 3.5 で述べる。

$$dch_i = \left(w_{i1} \quad w_{i2} \quad \cdots \quad w_{in} \right) \quad (1)$$

ここで、 w_{ij} はドキュメント D_i のキーワード K_j に対する重み ($-1 \leq w_{ij} \leq 1$) であり、その値が正のとき適合性、負のときは非適合性、ゼロのときはどちらでもないことを示す。 n は総キーワード数である。

散策空間内の全ドキュメントの特性を式 2 に示す行列によって表現する。これをドキュメント特性行列と呼ぶ。

$$DCH = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix} \quad (2)$$

ここで、 DCH の行要素が 1 つのドキュメントに対する特性を示す。 m は総ドキュメント数、 n は総キーワード数である。

3.2 読書特性

利用者は利用者識別情報(名前、年齢、所属など)とその読書特性によって表現される。読書特性はドキュメント特性と同様にベクトルモデルで表現し、読書特性ベクトルと呼ぶ。式 3 に任意の利用者の読書特性ベクトルを示す。読書特性ベクトルでのキーワードは、ドキュメント特性ベクトルと同一である。各重みの初期値は利用者自身が登録する。

$$RCH = \left(w_1 \quad w_2 \quad \cdots \quad w_n \right) \quad (3)$$

ここで、 w_i はキーワード K_i に対する重み ($-1 \leq w_i \leq 1$) であり、その値はドキュメント特性ベクトルの重みと

同様にキーワードへの適合性/非適合性を示す。 n は総キーワード数である。

3.3 散策特性

散策特性として、「散策中に利用者がどのドキュメントをどれくらい参照しているのか (参照履歴特性と呼ぶ)」と「どのドキュメントにどれくらい興味を持っているのか (興味度特性と呼ぶ)」で表す。

3.3.1 参照履歴特性

参照履歴特性は、利用者がドキュメントをどのような場所でどれくらい参照したかで定義する。本研究では、後述する「図書空間」、「個人用図書空間」および「個人書棚」での以下の参照行為を対象とする。

- (a) 図書空間/個人用図書空間でのドキュメントの参照
- (b) 個人書棚へのドキュメントの登録
- (c) 個人書棚でのドキュメントの参照

式4にドキュメント D_i に対する参照履歴特性を示す。

$$h_i = W_a \cdot n_{ai} + W_b \cdot n_{bi} + W_c \cdot n_{ci} \quad (4)$$

ここで、 n_{ai} , n_{bi} および n_{ci} は各行為が行われた回数である。ただし利用者がドキュメントを個人書棚から削除した場合には、その時点でそのドキュメントの n_{bi} は「0」とする。 W_a , W_b および W_c は各行為に対する重みである。同一のドキュメントを参照する場合、公共的な場所よりも個人的な場所の参照の方がより重要なものと仮定し、各重み間は $W_a \leq W_b \leq W_c$ とした ($W_a + W_b + W_c = 1$)。

式5に全ドキュメントに対する参照履歴特性を示す。これを参照履歴特性ベクトルと呼ぶ。

$$H = \begin{pmatrix} h_1 & h_2 & \cdots & h_m \end{pmatrix} \quad (5)$$

ここで、 m は総ドキュメント数である。

3.3.2 興味特性

興味特性は、利用者が参照した各ドキュメントに対する興味度で表現する。興味度は参照する度に変わる場合があると考え、ドキュメントを参照する度に興味度を入力し、その平均値を興味特性とする。式6にドキュメント D_j に対する興味特性を示す。

$$i_j = \frac{1}{n} \left(\sum_{k=1}^n \varepsilon_k \cdot r_k \right) \quad (6)$$

ここで、 n は入力回数、 r_k はドキュメント D_j に対し利用者が k 回目に入力した興味度 ($-1 \leq r_k \leq 1$)、 ε_k はそのときの興味係数 ($0 < \varepsilon_k \leq 1$) で、ドキュメントを参照した場所に依存したものである。ただし試作システムでは $\varepsilon_k = 1$ とし、興味度を同一に扱っている。興味度 r_k の値が正のときは興味の度合、負のときは非興味の度合、ゼロのときは興味があるかないか分

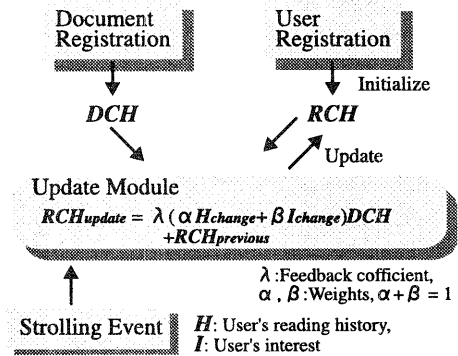


図2 読書特性の更新プロセス

Fig. 2 Process of Updating Reading Character.

からない/どちらでもないことを示す。ただし利用者が興味度を指定しない時には興味度に変化がないと考え、 $r_k = r_{k-1}$ とする。

式7に全ドキュメントに対する興味特性を式7に示す。これを興味特性ベクトルと呼ぶ。

$$I = \begin{pmatrix} i_1 & i_2 & \cdots & i_m \end{pmatrix} \quad (7)$$

ここで、 m は総ドキュメント数である。

3.4 読書特性の更新

予め利用者が自身の読書特性を正確に把握しているとは限らず、また散策を通して読書特性に変化が生じる場合も考えられる。例えばある特性を持ったドキュメントを頻繁に参照する場合、その利用者の読書特性はそのドキュメントの特性に近くなる。またその逆にほとんど参照しないドキュメントの特性とは異なったものとなると考えられる。そこで、本研究では散策特性の変化をドキュメント特性にマッピングし、それを読書特性にフィードバックすることで、散策過程を考慮した読書特性を求める。その手続きを式8に示す。

$$\begin{aligned} RCH_{update} &= \lambda \cdot (\alpha H_{change} + \beta I_{change}) DCH \\ &+ RCH_{previous} \end{aligned} \quad (8)$$

ここで、 RCH_{update} は更新した読書特性、 H_{change} は前回更新した時の参照履歴特性ベクトルとの差ベクトル、 I_{change} は前回更新した時の興味度特性ベクトルの差ベクトル、 $RCH_{previous}$ は更新前の読書特性である。 λ はフィードバック係数 ($\lambda > 0$) で、どれくらい散策特性をフィードバックさせるかを定めるものである。この値が大きいほど強く散策特性をフィードバックすることになる。 α と β は参照履歴特性と興味度特性どちらを重視するかを指定する係数である ($\alpha + \beta = 1$)。試作システムでは利用者がスライダを操作して興味特性を入力するため、参照特性よりも興味特性の方に散

策特性が反映されると考え、 $\alpha < \beta$ とした。図2に更新プロセスを示す。更新手続きは以下に行われる。

Step-1: 読書特性ベクトルの初期化

新規利用者は読書特性ベクトルのキーワードの重み w_i ($-1 \leq w_i \leq 1$) を登録する。

Step-2: 散策特性の獲得

利用者がドキュメントを参照すると、その参照履歴が後述のユーザデータベースに格納される。利用者がドキュメントに対し興味/非興味を持った場合、読書ブラウザに用意されているスライダで興味度を入力すると、その値がユーザデータベースに格納される。

Step-3: 読書特性ベクトルの更新

獲得した散策特性を入力とし、式8に基づいて読書特性ベクトルを更新する。その更新結果およびその履歴がユーザデータベースに格納される。

Step-4: 散策空間の更新

ユーザデータベースに格納された散策特性、読書特性の変化に応じて、後述の散策空間も更新される。詳細については4.5で述べる。

3.5 キーワード選定/重みの割り当て

ドキュメント特性ベクトルの作成は、ドキュメントのコンテンツの種類に依存する。例えばテキストデータの場合、形態素解析により出現頻度の上位のものをキーワードとして採用する方法が使われる。またイメージデータやビデオデータの場合、テキストデータの特性やイメージ内のコンテンツの空間的関係を使う方法などがある¹⁰⁾。しかしキーワードの自動抽出は必ずしも技術的に確定したものでなく、例えば出現語のうち否定の意味で用いられているといった構文的情報が必要になり、通常用いられているTF/IDF法のような重み付け法を拡張するなど、現在多くの研究がなされている。

本研究ではキーワード抽出は、ドキュメント登録時に人手によってキーワードを与える方式を採用する。キーワードの選定においてはキーワード辞書や各種のカテゴリ分類を用いる。重みの設定も同様に人手で行う。試作システムでは、キーワード選定に雑誌データに関しては情報処理学会で作成された「論文投稿のためのカテゴリ表」を参考にした。ただし雑誌データにあらかじめキーワードが付けられている場合にはそれを採用した。また映画データに関しては「ぴあシネマクラブ¹¹⁾」や映画データベース (<http://us.imdb.com/>) の映画分類を参考にした。ドキュメント特性ベクトルの各キーワードに対する重みは、ドキュメント登録時

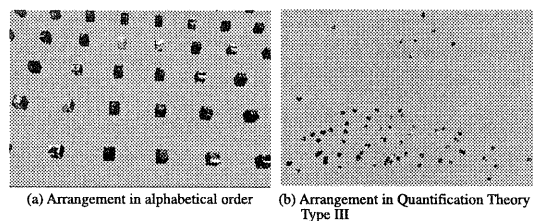


図3 映画データの図書空間の例

Fig. 3 An example of library space.

に著者の研究室の複数人によって決めた。読書特性ベクトルの各重みの初期値は、利用者自身が与えた。

本研究では、利用者の特性を読書特性ベクトル(キーワード空間上の1つのベクトル RCH) で表現し、散策対象のコンテンツの分野毎に読書特性ベクトルを用意する。通常、散策過程においては利用者の興味分野が散在すると考えられ、例えばエージェントを用いて利用者の複数の興味について機械学習により時間変化に追従する手法がある¹³⁾。この手法に比べ本手法では利用者の複数の興味分野を統合的に扱うことが困難であり、この手法の導入を含め、この点に関しては今後の課題である。

4. 本手法における散策空間

参照する状況に応じて、散策空間を「主に散策を目的とする空間(図書空間/個人用図書空間と呼ぶ)」と「参照したドキュメントに対し加工を行い、その結果を保存するための空間(個人書棚と呼ぶ)」に分ける。

4.1 図書空間

図書空間は、利用者が登録された全てのドキュメントを散策するための空間である。図書空間ではドキュメントを同一形状のオブジェクトで表現し、ドキュメント同士の類似性に基づいて配置する。配置位置は多変量解析手法の一つである数量化3類のカテゴリ散布によって求める*。図3に映画データの図書空間を示す。同図(a)は各オブジェクトをアルファベット順に配置した空間、同図(b)は数量化3類によって各オブジェクトを配置した空間である。(a)に比べて(b)の方が類似するドキュメントが物理的に近い場所に配置されるために、利用者が容易に類似したドキュメントを見つけることができる。

4.2 個人用図書空間

個人用図書空間は、利用者が各自の読書特性や散策特性の観点から図書空間を散策するための空間である。

* ドキュメント特性ベクトル dch の要素 w の大きさは、 $-1 \leq w \leq 1$ の範囲である。数量化3類の計算過程では、 $-1 \leq w \leq 0$ を「0」、 $0 < w \leq 1$ を「1」とした。

個人用図書空間でもオブジェクトは図書空間と同様にドキュメント間の類似性に基づいて配置される。ただし対象とする特性の大きさに応じて、各オブジェクトの形状が異なる。個人用図書空間としては以下の5種類の空間がある。オブジェクトの形状に関しては4.4で述べる。

- 類似特性空間：ドキュメント特性と読書特性との類似特性の観点による空間*
- 参照履歴特性空間：参照履歴特性の観点による空間
- 興味特性空間：興味特性の観点による空間
- 複合特性空間：類似特性、参照履歴特性および興味特性の観点による空間
- 平均特性空間：複数利用者の上記特性の平均値の観点による空間

これらの空間を散策することで、図書空間では分からなかった各利用者の興味のあるようなオブジェクトを簡単に把握することが可能となる。個人用図書空間では、利用者が設定するしきい値によって、各特性空間で表示するオブジェクトにフィルタリングを行う。また利用者が各特性を反映せずに、さらに表示選択を行わずにドキュメント群を散策する場合には、図書空間を利用する。

4.3 個人書棚

利用者が図書空間／個人用図書空間を散策した際に、興味のあるドキュメントを登録し、そのドキュメントに注記、メモ書きやCut-and-Pasteによるファイリングなど様々な加工を行うための空間である。書棚に登録されたドキュメントは、参照時間順または個人用図書空間と同様の特性を用いてソーティングされ表示される。個人書棚へのドキュメントの登録・削除には、Webブラウザの「ブックマーク」や「お気に入り」と同様にその索引情報(後述のドキュメントサーバへのリンク情報)だけを扱う。またドキュメントへの加工操作は、ドキュメントがHTML文書の場合、XML(eXtensible Markup Language)¹²⁾のスタイルシートや双方向ポインタ(XPointer)を利用する。すなわち各利用者が各ドキュメントに対しスタイルシートを作成しXPointerを付加することで、オリジナルのドキュメントに手を加えることなく加工操作を実現する。ただし試作システムでは加工機能に関して実現を試みている段階である。これらの登録・削除履歴やスタイルシートは、後述の仮想図書館サーバ内のユーザデータベースで各利用者ごとに管理され、各利用者の個人書棚を構築する際に使われる。

4.4 散策空間でのドキュメント表現

ドキュメント表現には本メタファを利用し、利用者特性やドキュメント特性を「本」の形状に反映させ、「本」を眺めることで、どのような情報であるかを把握できるようにする。その形態としては「3D図書オブジェクト」と「仮想図書」がある。前者は散策空間にドキュメントが配置されるときに使われる。後者はドキュメントのコンテンツの参照や加工のために使われる。図4に3D図書オブジェクトと仮想図書を示す。

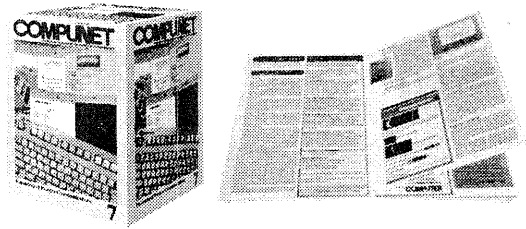


図4 3D図書オブジェクトと仮想図書
Fig. 4 3D book object and virtual book.

4.4.1 3D図書オブジェクト

3D図書オブジェクトの形状は直方体で、その表面にはドキュメントの表紙イメージデータをテキストチャマッピングする。視点位置によって前方のオブジェクトが壁になり後方のオブジェクトが見えない場合があるため、半透明イメージを用いる。散策時には視点からの距離に応じてオブジェクトの透明度を変化させる機能がある(これをLOD:Level of Detail機能と呼ぶ)。また視覚的な内容把握が適していないドキュメント(例えば文章)に対応して補助的にオブジェクトを指定するとそのタイトル情報を表示する。

図書空間では全ての3D図書オブジェクトは同一サイズである。一方、個人用図書空間の各特性空間では、対応する特性を3D図書オブジェクトの「サイズ」と「透明度」に反映させる。すなわち対応する特性が大きいほど、オブジェクトのサイズが大きく透明度が低くなる。オブジェクトサイズの大小関係によって特性の大小関係を把握することが困難な場合にも、透明度の変化を利用することで容易に他のオブジェクトの特性との相対的關係を把握できる。例えば図5に類似特性空間を示す。同図で視覚的にはObject-AとObject-Bのサイズに差はないが、透明度によってObject-BがObject-Aよりも利用者との類似性が高いことが分かる。

また複合特性空間での3D図書オブジェクトは、直方体の各1辺が各々類似、参照および興味特性に対応

* 付録A.1に両ベクトル間の類似度の定義を示す。

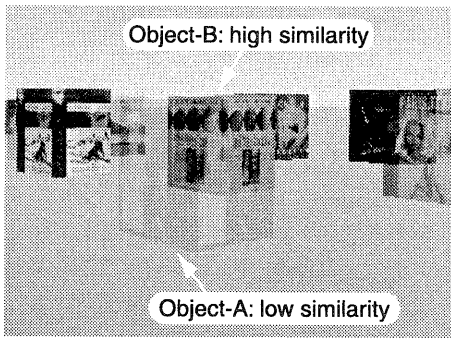


図5 類似特性を反映した3D図書オブジェクト
Fig. 5 3D book object reflected similarity

している。視点を移動させてオブジェクトを眺めることで類似度、参照履歴および興味度を把握できる。さらに、ある軸方向からオブジェクトを眺めると、オブジェクト表面に張り付けたドキュメントの表紙イメージの歪みによって2つの特性間のバランスが分かる。

4.4.2 仮想図書

各ドキュメントのコンテンツ提示に、実際の本のようにばらばらとページをめくる感覚で参照するタイプと、詳細に中身を参照するタイプを導入する。前者はWebBook⁵⁾と同様にドキュメントの各ページのイメージデータを仮想的な本に整形し提示する。後者は文書の場合にはテキスト文書やPDF (Portable Document Format) 文書、イメージデータやビデオデータの場合には該当するイメージや映像を提示する。

4.5 散策空間の更新

散策空間の更新は以下の場合に行われる。

- 散策特性／読書特性が変化したとき
- 新規のドキュメントが登録されたとき

前者では、その変化に応じてシステムが逐次的に更新操作を行う設定、および利用者の更新指示に応じて更新が行われる設定がある。どちらの設定でも個人用図書空間と個人書棚にある3D図書オブジェクトのサイズと透明度が更新される。後者では、最初に新規ドキュメントを非透明かつ利用者との類似性に応じたサイズのオブジェクトで表示する。そして既存ドキュメントのうち新規ドキュメントと最も類似性をもつオブジェクトの近くに配置する。その後、利用者の視点を新規オブジェクトに合わせる。ただし利用者が既存のオブジェクトの位置を頼り散策を行う場合があるため、図書空間／個人図書空間の全オブジェクトの再配置は利用者からの指示によって行う。ここでのオブジェクト間の類似性は付録A.1に述べる定義である。

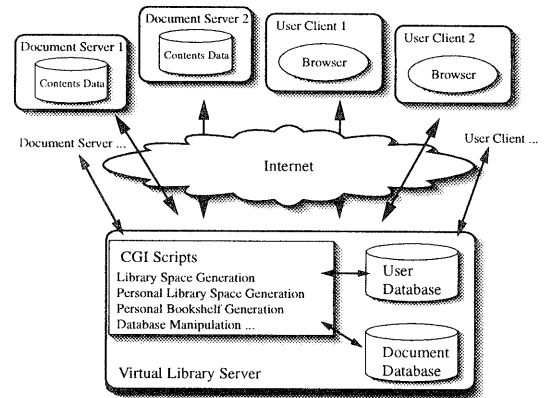


図6 システム構成
Fig. 6 System Structure

5. システムアーキテクチャ

図1に示した試作システムは、図書空間／個人用図書空間部分、個人書棚、ドキュメントを表示する部分(同図ではWeb Page)やコントロール部からなる。現在、コントロール部には、個人書棚ヘドキュメントを登録・削除を行うためのプルダウンメニュー、表示する図書空間／個人図書空間の切替えのためのプルダウンメニュー、タイトルによる検索部、数量化3類による類似性に基づく配置、LOD機能のON/OFFボタン、新規ドキュメント登録を常時監視する機能のON/OFF用のボタン、読書特性を随時変更する機能のON/OFF用のボタンがある。

試作システムはインターネット上の情報リソースへの対応からWWW環境を利用した。図6にシステム構成を示す。各要素について以下に述べる。

仮想図書館サーバはドキュメントを管理するドキュメントデータベース、利用者情報を管理するユーザデータベース、それらのデータベースから図書空間および個人用図書空間を生成するためのスクリプト群からなる。ドキュメントデータベースには、ドキュメントのコンテンツデータの格納先情報(URL情報)やドキュメント特性が格納される。ユーザデータベースには各ユーザの読書特性、散策特性および個人書棚での履歴が格納される。仮想図書館サーバおよびドキュメントサーバには、WWWサーバソフトウェア(Apache 1.3.9)を利用した。ユーザクライアントからデータベースへの読み書きには、PerlによるCGIスクリプトで実装した。またJava AppletからVRML2.0 (Virtual Reality Modeling Language)¹⁴⁾で記述した仮想空間を動的に操作するために、EAI (External Authoring Interface)¹⁵⁾を利用した。

ユーザクライアントは仮想図書館サーバにある Java Applet をダウンロードすることで機能する。ユーザクライアントには Windows 98 および SGI IRIX6.5 上 Netscape Communicator 4.7 に VRML2.0 用のプラグインソフトウェア (Cosmo Player 2.1) をプラグインしたものを用いた。

ドキュメントサーバにはドキュメントのコンテンツが格納される。各コンテンツデータは URL 情報によって参照される。

6. システムの動作

試作システムでは、ドキュメントとして雑誌データと映画データを利用した。雑誌データとしては 50 タイトル、15 キーワード (マルチメディア, ソフトウェア, コンピュータグラフィックス, ...) を用いた。映画データとしては 100 タイトル、20 キーワード (コメディ, SF, アクション, ...) を用いた。ここで雑誌データは本研究室で購入した学術雑誌をスキャナーで生成したものであり、ドキュメントサーバは外部のネットワークとは物理的に孤立してある。また映画データは対象とした全映画タイトルの映像データが利用できなかったため、インターネット上の映画データベース (<http://us.imdb.com/>) の Web ページを代用した。この Web ページには該当する映画の監督、俳優などの情報の他に、Web ページ利用者が投票して決めたランキング情報などが含まれている。

各特性を計算するため各パラメータ値は、式 4 では $W_a = 0.1$, $W_b = 0.3$ および $W_c = 0.6$, 式 6 では $\epsilon = 1.0$, 式 8 では $\lambda = 1.0$, $\alpha = 0.3$ および $\beta = 0.7$ とした。これらの値は必ずしも最適値ではなく、実験的なものである。

6.1 読書特性更新に関する実験

5 人の被験者に対し、雑誌データと映画データを用いて利用者実験を行った。利用者の散策特性がどのように読書特性に影響するのかを調べるために、利用者は図書空間だけ散策した。実験後の被験者のコメントから試作システムの散策機能は文字情報を中心とした雑誌データよりも画像情報を多く含んだ映画データに適していることがわかった。それは、雑誌データでは散策中に 3D 図書オブジェクトでタイトル情報は分かるが、そのテキストチャイメージでは内容把握が容易でない。一方映画データの場合はそのテキストチャイメージである程度内容が把握できるためである。雑誌データのように文字情報を中心としたコンテンツへの対策は今後の課題である。そこで以下の利用分析には映画データの散策結果を用いる。

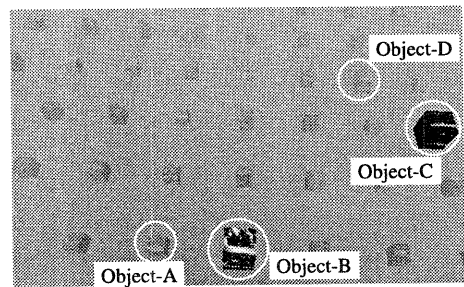


図 7 散策後の参照履歴特性空間

Fig. 7 An example of history space after strolling.

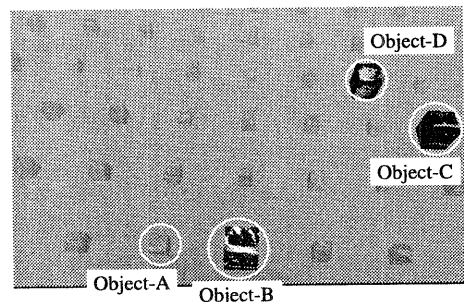


図 8 散策後の興味特性空間

Fig. 8 An example of interest space after strolling.

6.1.1 実行例

ひとりの被験者の散策過程の分析結果を以下に示す。被験者が予め入力した読書特性では、「コメディ」、「ファミリー」および「ロマンス」への重みが大きく (重みが 1), 「SF」や「アクション」には重みが小さかった (重みが 0.0)。散策後の参照履歴特性空間と興味特性空間を図 7, 図 8 に示す*。同図で Object-A は題名「The Cure」、特性「ファミリー, ドラマ, キッズ」、Object-B は題名「Apollo 13」、特性「SF, ドラマ, アクション」、Object-C は題名「Independence Day」、特性「SF, アクション, 戦争」および Object-D は題名「Big」、特性「コメディ, SF, ドラマ」である。図 7 から Object-B と C に対する参照履歴特性が他のものに比べると強いことが分かる。それを裏付けるように Object-B と C を個人書棚に登録していた。また Object-A と D は他のドキュメントとほぼ同一の参照特性であった。図 8 から興味特性が Object-C が最も高く、B, D, A の順に高いことがわかる。Object-D は、図 7 と図 8 で最も差があるが、これは被験者は個人書棚には登録しなかったが、興味を強く持ったためである。このように

* これらの実行画面は、簡略に説明するためオブジェクトを格子配置したものである。また図 9(a), (b) も同様である。

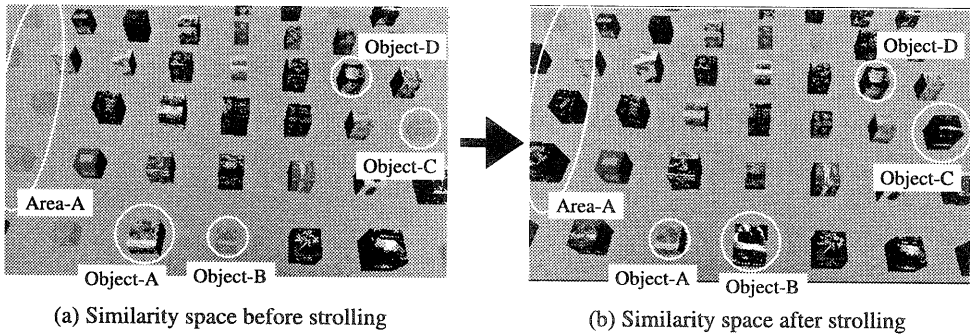


図9 読書特性更新前後の類似特性空間

Fig. 9 An example of similarity space before/after updating reader character.

参照履歴特性だけでは必ずしも利用者が興味を持ったものであるとは言えない場合がある。

散策特性を用いて更新した読書特性とドキュメントの類似度は、Object-A が 0.9 から 0.844, Object-B が 0.825 から 0.931, Object-C が 0.825 から 0.911, Object-D が 0.9 から 0.910 と変化した。この変化を表示したものを図 9(a), (b) に示す。Object-B と C は同図 (a) ではオブジェクトのサイズが小さくそして透明度が高かった。同図 (b) ではサイズが大きく、透明度が低くなった。これは SF やアクションに関して読書特性に変化があったことを示している。また Object-A については同図では (b) の方がサイズが小さくなっている。これはファミリーに関して読者特性に変化があったことが一つの要因と考えられる。このように、実際の散策過程では被験者が予め入力した読書特性と異なったドキュメントに対して興味を示したことが分かる。

同図 (a) と (b) で Area-A にあるドキュメント群に変化がある。ドキュメント群には、題名「Mask」で特性「コメディ, SF, アクション」、題名「Heat」で特性「ドラマ, サスペンス, アクション」、題名「ジェラシックパーク」で特性「SF, サスペンス, アクション」や題名「From Dusk Till Dawn」で特性「ドラマ, サスペンス, アクション」が含まれ、散策中に被験者はこれらのドキュメントを参照していない。しかし、読書特性の更新により SF やアクションの重みが増したことで、それらのドキュメントとの類似性が高くなり、結果的に Area-A にあるドキュメント群がはっきりと現れた。このように本手法では、利用者が参照していない未知のドキュメントを有用なドキュメントとして新たに提示可能である。

散策空間について以下のような有用性が明らかになった。

- 散策過程で利用者が興味を示したドキュメントが

表1 読書特性と散策空間内のドキュメント特性との類似度
Table 1 Results of change of similarity.

実験	最小類似度	最大類似度	平均類似度	分散 ($\times 10^{-4}$)	レンジ
Case-0	0.775	0.950	0.8727	3.84 ²	0.175
Case-1	0.786	0.937	0.8778	2.62 ²	0.151
Case-2	0.783	0.934	0.8759	2.57 ²	0.151
Case-3	0.782	0.932	0.8749	2.56 ²	0.150
Case-4	0.781	0.931	0.8739	2.57 ²	0.150
Case-5	0.780	0.929	0.8727	2.60 ²	0.149

最大小値、平均値および分散は各々小数点以下第 4,5,6 位で四捨五入をした。

明らかになる。すなわち参照履歴特性空間や興味特性空間を眺めることで、散策特性が容易に把握できる。

- 新規ドキュメント (未参照ドキュメント) に対する推薦機能として利用できる。

6.1.2 読書特性更新に関する評価

本研究で導入した読書特性更新手続き (式 8) では、参照履歴特性と興味特性のどちらを重視するかによって更新結果が異なる。そこで α と β の値を変化させ、どのように読書特性が更新されるかを調べる。そのため以下の 6 つのケースについて調べた。各ケースに関して表 1 に読書特性とドキュメント特性との類似度の最小値、最大値、平均値、分散およびレンジ (最大値と最小値の差) を示す。各値は全被験者平均値である。
Case-0 読書特性を更新しない場合 ($\alpha = 0.0, \beta = 0.0$)

Case-1 参照履歴だけを反映させる場合 ($\alpha = 1.0, \beta = 0.0$)

Case-2 参照履歴を重視する場合 ($\alpha = 0.7, \beta = 0.3$)

Case-3 参照履歴と興味特性を同等に反映させる場合 ($\alpha = 0.5, \beta = 0.5$)

Case-4 興味特性を重視する場合 ($\alpha = 0.3, \beta = 0.7$)

表 2 類似度特性に関するドキュメントの分布

Table 2 Results of document distribution from a view of similarity.

実験	範囲 1 (個数)	範囲 2 (個数)	範囲 3 (個数)
Case-0	20	65	15
Case-1	15	71	14
Case-2	17	71	12
Case-3	17	71	12
Case-4	17	71	12
Case-5	17	70	13

s : 類似度, \bar{S} : 平均類似度および σ^2 : 分散. 範囲 1: $s < \bar{S} - \sigma$, 範囲 2: $\bar{S} - \sigma \leq s \leq \bar{S} + \sigma$ および 範囲 3: $\bar{S} + \sigma < s$.

Case-5 興味特性をだけを反映させる場合 ($\alpha = 0.0$, $\beta = 1.0$)

表 1 より Case-1~5 に関しては α と β によって各項目での差はほとんどない. これは, 実験中, 被験者が参照履歴特性の高いオブジェクトに高い興味特性を入力する傾向で, 結果的に α と β の値によって各項目に差が現れなかった. 一方更新を行わない場合 (Case-0) に比べ, その以外の場合では平均類似度が若干高くなり, また分散およびレンジに関しては大きく減少している. 平均類似度の変化がわずかなため, 更新した読書特性と全ドキュメントの特性が類似する傾向にあると必ずしもいえない. しかし分散およびレンジが減少していることより, 平均類似度付近にあるドキュメント数が増加したと考えられる. そこで平均類似度 \pm 標準偏差の範囲にいくつドキュメントがあるかを調べ, その結果を表 2 に示す. 同表より平均類似度付近にあるドキュメント数が増加したことが分かる.

表 1, 表 2 より, 空間の全ドキュメントに対する平均類似度はほとんど変化しないが, 平均付近にあるドキュメントが増加したことより, 散策を行うことで散策空間内のドキュメントの平均的なドキュメント特性に近づくことを示している. これは, 表 1 の Case-0 より元々被験者の読書特性と映画データとが高い類似性があり, 先の述べた被験者の参照履歴特性と興味特性の関連から, 読書更新による効果が明確にならなかった. そのため, 被験者の特性や使用するコンテンツデータの構成を含め, より厳密な読書特性の更新プロセスに関する評価を行う必要がある. このことは今後の課題である.

6.2 個人用図書空間に関する実験

個人用図書空間を散策することで, (1)「利用者がより興味度の高いドキュメントを参照することができるか」, (2)「利用者があまり興味のないドキュメントを参照しないですむか」の 2 点を検証するために, 新た

表 3 図書空間と個人用図書空間で参照したドキュメント種類数と興味度に関する結果

Table 3 Results of the number of kinds of documents that each subject referred to in library space and personal library space, and the average of interest degree

被験者	図書空間		個人用図書空間	
	種類数	平均興味度	種類数	平均興味度
A	21	0.32	26	0.35
B	26	0.42	21	0.72
C	10	0.80	4	0.75
D	40	0.53	23	0.58
E	7	0.14	6	0.53
F	24	0.44	27	0.59
G	70	0.09	32	0.73
H	16	0.71	17	0.77
I	24	0.40	11	0.72
総数	238	—	167	—
総平均	—	0.428	—	0.638

小数点以下第 3 位で四捨五入をした.

表 4 図書空間と個人用図書空間で非重複参照のドキュメント種類数と興味度に関する結果

Table 4 Results of the number of kinds of documents that each subjects did not refer to in both library space and personal library space, and the average of interest degree.

被験者	図書空間		個人用図書空間	
	種類数	平均興味度	種類数	平均興味度
A	5	0	10	0.37
B	18	0.22	13	0.67
C	8	0.79	2	0.55
D	23	0.40	6	0.38
E	4	-0.18	3	0.53
F	19	0.42	22	0.62
G	55	-0.06	17	0.55
H	11	0.64	12	0.70
I	16	0.13	3	0.30
総数	159	—	88	—
総平均	—	0.262	—	0.519

小数点以下第 3 位で四捨五入をした.

に 9 人の被験者に対し被験者実験を行った. 被験者は図書空間を散策し, 興味のあるようなドキュメントを参照し興味度を入力する. この作業を被験者自身が興味のあるドキュメントを探し終ると判断するまで繰り返す. 個人用図書空間に対しても同様のことを行った.

各空間において参照したドキュメントの種類数と入力した興味度を表 3 に示す. 同表より, 被験者の多くが個人用図書空間で参照したドキュメント数が少なく, その平均興味度が高い. また総平均興味度も個人用図書空間の方が高い. このことは, 被験者が個人用図書空間でより興味度の高いドキュメントを参照する傾向にあることを示す. したがって (1) の点を確認できた.

(2)を検証するために、ここでは図書空間と個人用図書空間で重複して参照したドキュメントは被験者にとって興味があるものとし、重複しないドキュメントに関して、その種類数とその平均興味度を表4に示す。同表より、被験者EとGの平均興味度は図書空間では負であるが、個人用図書空間ではその総平均興味度よりも高い。また総平均興味度も個人用図書空間の方が高い。このことは、個人用図書空間を散策する方があまり興味のないドキュメントを参照することを回避できたことを示す。したがって(2)の点を確認できた。

7. おわりに

本論文では、利用者が散策時にオブジェクトが示しているコンテンツと、そのドキュメント特性と読書特性の関係を容易に把握するための手法を提案し、その手法に基づいてシステムを実装した。すなわち、それらの特性の視点から散策空間内に配置した3D図書オブジェクトの形状を変化させた。また散策特性によって読書特性を更新し、それに基づいて散策空間を更新した。このようにして、利用者に適合した散策空間を提供した。その結果、個人用図書空間を眺めることで、ドキュメント特性や利用者特性が容易に把握でき、新規ドキュメントの推薦に利用できた。また利用者はより多くの興味のあるドキュメントを参照することができた。逆にあまり興味のないドキュメントへの参照を減らすことができた。

今後の課題としては本手法のより厳密な評価が挙げられる。例えば、より規模の大きな散策空間に適用し、スケーラビリティの評価を行う。また各パラメータを変化させ、読書特性の更新手法の妥当性の検証を行う。さらに、本研究では散策によりドキュメント特性が読書特性に反映するように扱っているが、逆に読書特性がドキュメント特性に影響を及ぼすと考えることもできる。このように読書特性とドキュメント特性の相互作用などの検証実験を行う。

謝 辞

映画データを快く御提供して下さった電気通信大学大学院情報システム学研究科の小池英樹助教授と折茂栄美子さんに感謝します。なお、この研究は栢森情報科学財団の助成を受けて遂行された。

参 考 文 献

- 箱崎勝也, 金井秀明, 石川克則 他, “個人利用に適合した仮想図書館の構想,” 電子情報通信学会論文誌, Vol.J81-D-II, No.5, pp.1025-1034, 1998.
- T.W. Yan and H. Gracia-Molina, “SIFT - A Tool for Wide-Area Information Dissemination,” USENIX Technical Conference, pp.177-186, 1995.
- T.Kamba, B.Krishna, and M.C.Albers, “KrakatoaChronicle: an interactive personalized newspaper on the Web,” *Proc.of 4th International WWW Conference*, pp.159-170, 1995.
- K.Lang, “Learning to Filter Netnews,” *Proc.of 12th International Conference Machine Learning (ICML '95)*, pp.331-339, 1995.
- Stuart k.Card, George G.Robertson and William York, “The WebBook and the Web Forager: An INformation Workspace for the World-Wide Web,” *Proc. of Human factors in computing systems '96 (CHI '96)*, pp.111-117, ACM Press, 1996.
- G. Robertson, M. Czerwinski and et al, “Data Mountain: Using Spatial Memory for Document Management,” *Proc.of ACM symposium on User Interface Software and Technology (UIST '98)*, pp.153-162, ACM Press, 1998.
- 田中豊, 脇本和昌, 多変量統計解析法, 現代数学社, 1998.
- Emiko Orimo and Hideki Koike, “ZASH: A Browsing System for Multi-Dimensional Data,” *Proc. of IEEE Symposium on Visual Languages '99 (VL '99)*, pp.288-295, IEEE Computer Society, 1999.
- Junichi Tatemura, “Visualizing Documents Space by Force-directed Dynamic Layout,” *Proc. of IEEE Symposium on Visual Languages '97 (VL '97)*, pp.119-120, IEEE Computer Society, 1997.
- Alberto del.Bimbo, “Visual Information Retrieval,” Morgan Kaufmann, 1999.
- “びあシネマクラブ,” ぴあ出版, 1999.
- Goldfarb, C.F, Prescond, P., “The XML Handbook '98,” PH-PTR, 1998.
- A. Moukas and G. Zacharia, “Evolving a multi-agent information filtering solution in Amalthea,” *Proc. of first international conference on Autonomous agents (Agents '97)*, pp.394-403, ACM press, 1997.
- J. Hartman, J. Wernecke, “The VRML 2.0 Handbook-Building Moving Worlds on the Web,” Addison-Wesley, 1996.
- Chris Marrin, “Proposal for a VRML 2.0 Informative Annex External Authoring Interface Reference,” <http://vrml.sgi.com/moving-worlds/spec/ExternalInterface.html>, 1997.

付 録

A.1 類似特性の定義

ベクトル間の距離概念(非類似度)を用いて,ベクトル間の類似度を定義する.従来より類似度の定義には,ベクトル間のなす角度の余弦(余弦尺度)が利用されている.この方法はベクトルの各成分の大きさを考慮しない場合には有効であるが,スカラー倍されたベクトルに対しても同一の類似度が求まる.本研究では各成分の大きさ(各属性の重み)を議論するため,ベクトルの各成分の大きさが反映される方法で類似度を定義する.そこで L_1 距離(マンハッタン距離)を用いる.すなわちベクトル $A:(w_{a1}, w_{a2}, \dots, w_{an})$ とベクトル $B:(w_{b1}, w_{b2}, \dots, w_{bn})$ との距離 d の定義を以下に示す.

$$d = \sum_{j=1}^n |w_{aj} - w_{bj}| \quad (9)$$

ここで各ベクトルの要素 w の大きさは, $-1 \leq w \leq 1$ とする. n はベクトルの要素数である.

類似度をその非類似度と非類似度の最大値との差で定義する.そして正規化すると以下ようになる.

$$\text{Similarity} = 1 - d/2n \quad (10)$$

この類似度定義を読書特性とドキュメント特性の類似度およびドキュメント特性間の類似度に利用した.

(平成 1999 年 12 月 20 日受付)

(平成 2000 年 3 月 25 日採録)

(担当編集委員 江口浩二)

金井 秀明 (正会員)



1969年生.1993年電気通信大学大学院博士前期課程修了.1996年同博士後期課程単位取得退学.同年,同大学院情報システム学研究科助手,現在に至る.工学博士.仮想図書館等の情報システムの研究に従事.情報検索システム,情報視覚化に興味を持つ.情報処理学会,IEEE,ACM各会員.

箱崎 勝也 (正会員)



1941年生.1963年九州大学工学部電子学科卒業.同年日本電気入社.中央研究所,ソフトウェア開発グループにおいて,システム性能評価,コンピュータアーキテクチャ,OS,ネットワークの相互接続性などの研究開発に従事.1994年電気通信大学大学院情報システム学研究科教授.工学博士.分散システム運用技術,マルチメディア利用技術,情報システムアーキテクチャなどの研究に従事.情報処理学会論文賞受賞(1982).情報処理学会,電子情報通信学会,IEEE,ACM各会員.

石川 克則



1971年生.1998年電気通信大学大学院情報システム学研究科修了.同年,凸版印刷入社,現在に至る.インタラクティブメディアを用いた情報視覚化に興味を持つ.