

ユーザの視聴傾向に基づき個人化した Web ページの動的作成

福 村 真 哉[†] 中 野 賢^{††} 春 本 要[☆]
下 條 真 司[☆] 西 尾 章 治 郎[†]

近年、デジタル化されたテキストや画像、音声など、様々なデータが WWW (World Wide Web) 上で広く提供され、多くのインターネットユーザによって活用されている。しかし、Web サイトの多くは、ハイパーリンクによって Web ページの構成が固定されているため、ユーザが欲する情報を容易に見つけるのは困難な場合がある。この問題に対し、筆者らは、新たな Web サイト管理モデルと、ユーザの Web サイトにおける視聴履歴を利用した Web ページの個人化手法を提案している。本論文では、この個人化手法において、Web ページを構成する要素の選択方法と、ユーザが容易に情報を発見できる Web ページの作成方法を詳しく述べる。

Dynamic Production of Personalized Web Pages Based on Browsing Pattern

SHINYA FUKUMURA,[†] TADASHI NAKANO,^{††} KANAME HARUMOTO,[☆]
SHINJI SHIMOJO[☆] and SHOJIRO NISHIO[†]

Recently, the tremendous amount of information is digitalized and becoming available on the World Wide Web. These digital contents are nowadays easily accessible to a lot of Internet users. However, most web sites today are comprised of a large number of web pages and hyperlinks, often making it difficult for visitors to browse through and to find their desired information quickly. To address the problem, we are proposing a newly web site management model and its personalization method by utilizing content browsing history. In this paper, we describe details about how the system chooses components to be displayed and how the system constructs personalized web pages from those components.

1. はじめに

近年、デジタル化されたテキストや画像、音声など、様々なデータが WWW (World Wide Web) 上で広く提供され、多くのインターネットユーザによって活用されている。例えば、美術館や博物館に展示されている作品およびその解説文は、画像データとテキストデータとしてデジタル化され、WWW 上で公開されている。ユーザは例えば、お気に入りの作品を、パソコンや携帯電話、PDA (Personal Digital Assistants) など、Web ブラウザを備えた端末を用いて、世界中のあらゆる場所から閲覧できる。WWW における情報提供は、多くのユーザにとって有用である。

しかし、Web サイトの多くは、現在、多数の Web ページとハイパーリンクによって固定された構成によって構築されている。そのため、美術館のサイトを例に例えと、美術作品の画像だけを見たいユーザや、画像と解説を同時に見たいユーザ、それぞれの要求に対して、提示する情報を変化させるなど、柔軟に対応するのは困難である。例えば、ユーザは興味のある情報や欲している情報をすぐに見つけることができず、情報を見つけるまでに何度もリンクを辿り、同じ行動を繰り返し、意味のない Web ページを閲覧する可能性がある。

この問題を解決するためには、Web サイトに含まれる情報の中から、ユーザの嗜好にあった情報を選び出し、それらを組み合わせ、ユーザに適切に提示することが重要である。そこで、筆者らは、情報の最小単位としてのコンポーネントおよびコンポーネントを関連付けるリレーション、さらにリレーションの意味と重みという概念からなる新たな Web サイトモデルを提案し、コンポーネントとリレーションに基づいて、ユーザに対して動的に個人化した Web ページを提示するシステムを提案している¹⁹⁾。

本論文では、システムにおけるユーザに適した情報の抽出部分および抽出した情報から Web ページを作成する部分に関する詳細を述べる。情報の抽出には、Web サイト作成者が付加したリレーションの重みと、ユーザの嗜好である視聴傾向を活用する。重みは、リレーションで関連付けられた二つのコンポーネントを Web サイト作成者が同時にユーザに提示したいとき、大きくなる。視聴傾向は、ユーザが過去に辿ったリレーションの意味列である視聴履歴から、データマイニング技術によって抽出する。Web ページの作成に関しては、効果的な提示のために、リレーションの意味を考慮して、コンポーネントの配置やユーザが辿りやすいハイパーリンクを実現する。

本論文の構成は次の通りである。まず、2 章で WWW における個人化に関する関連研究を紹介する。次に、3 章で筆者らが提案するシステムの概要を述べる。視聴履歴に基づく Web ページの動的作成手法は 4 章で提案する。5 章で提案を基に実装したシステムについて述べ、その後、今後の研究を 6 章で述べる。最後に 7 章で本研究をまとめる。

[†] 大阪大学大学院情報科学研究科 マルチメディア工学専攻
Department of Multimedia Engineering, Graduate School of
Information Science and Technology, Osaka University

[☆] 大阪大学 サイバーメディアセンター
Presently with Cybermedia Center, Osaka University

^{††} カリフォルニア大学アーバイン校
Department of Information and Computer Science, University
of California, Irvine

2. 関連研究

本章では、Web ページの個人化や提示方法の観点から本研究の関連研究を示す。

まず、ユーザの興味に基づく Web ページ推薦システムとして、Chi らの研究⁴⁾ や Zhu らの研究¹⁰⁾ がある。前者は、ユーザの欲する情報と、ユーザが閲覧している Web ページとの類似性を評価して、ユーザが次に訪れる可能性のある Web ページを予測する手法である。この手法では、Web ページ中のハイパーリンクとハイパーリンクを辿った先の Web ページに含まれる単語に関して、キーワードで表現されたユーザの欲する情報に対する適合度合いを評価する。評価結果は、ユーザがこの先訪れる可能性のある各 Web ページの重要度となる。この手法では、例えば「ここをクリックする」といった、文章として特に意味のないハイパーリンクに関しては、適当な結果が得られない。また、ハイパーリンクを辿った先にある Web ページの内容も、ユーザの興味を反映するとは限らない。後者は、前者と似ているが、検索エンジンの検索結果や HTML のタグ情報、メタ情報を利用して信頼度を上げている。しかし、Web サイト作成者の意図は反映しておらず、予測結果によっては、Web サイト作成者の見せたい情報がユーザに無視されるおそれもある。

清光、竹内らの研究^{17),18)} では、ユーザが過去に閲覧した Web ページや、辿ったハイパーリンクの情報に基づき、Web ページに提示する情報や Web ページ間の構造を動的に変化させる方法について議論している。これらの研究は、あらかじめ存在する Web ページを推薦、変更する点で本研究とは異なる。また、本研究は、情報を推薦するシステムではなく、ユーザの Web サイトにおける閲覧作業を容易かつ効果的に支援するシステムを構築する点においても、目的が異なる。

Web サイトのナビゲーションに関する研究として、Metro⁵⁾ や CAST²⁾、MAPA³⁾ といったプロジェクトがある。Metro は個々の Web ページを地下鉄の駅、ハイパーリンクをレールとみなすことで、ユーザに対して、Web サイトの構成を地下鉄の路線図に見立てて提示する。CAST は、Web サイトをツリー構造で表現したサイト案内ツールを Web ページの一部に提示するシステムであり、Web ページに注釈をつける機能も備えている。また、MAPA は、Web ページ間の階層情報と Web ページのアクセス権、Web ページのカテゴリを次元とする三次元空間に Web ページを配置することで、Web サイト全体の把握を容易にする。これらの研究は、ユーザにとって Web サイトを包括的に把握でき、欲する情報を見つけやすい利点があるため、本研究でもコンポーネント間の関係を把握する方法として利用できる。ただし、Web ページ数が増大すると、個々の Web ページの視認が困難になる問題がある。

Web ページの効果的な提示に関する研究として、三浦らはハイパーリンクの参照先を元の文書に埋め込む手法⁶⁾ を提案している。これは、ハイパーリンクで参照先に移動する手法とは異なり、クリックした位置に参照先の Web ページを展開する手法である。この手法を用いれば、一つの Web ページで、あらゆる情報の閲覧が可能となる。しかし、提示する情報量が増大すると、内容の把握が困難となり、重要な情報を見失うおそれがある。また、ブラウザ画面を上下左右にスクロールする作業など、ユーザの負担が大きくなる。

YoriDoriWeb⁷⁾ は、あらゆる Web ページの中で、ユーザ自身にとって必要な部分を集め、一つの画面にまとめ、一覧できるシステムである。しかし、画面に表示される Web ページの内容は、元々存在する Web ページに依存して変化せず、ハイパーリンク構造も変わらない。本研究では、ハイパーリンクを辿った先も動的にレイアウトを行うシステムを開発する。

また、Yahoo!をはじめとする各ポータルサイトは、個人化した Web ページを動的に生成するサービス¹⁶⁾ を実現している。これらのサービスは、ユーザの設定した趣味や Web ページレイアウトなどにしたがって、Web ページを作成、提供する。しかし、設定において、提示する情報の選択などが行える反面、設定作業がユーザの負担となる。また、Web ページの作成においては、例えばニュースなど、他のサイトから情報を直接切り出し、設定内容に合致するものをそのまま提示するだけであり、Web サイト全体に対して個人化を行うわけではない。

3. ユーザの視聴履歴に基づく個人化システム

Web サイトの閲覧における問題は、ユーザの多様な閲覧目的や嗜好に対し、それらを反映した情報を提示できないことである。そのため、ユーザは、興味のある情報や欲する情報に辿り着くまでに、何度もハイパーリンクをクリックしたり、不要な情報を閲覧する可能性がある。筆者らは、この問題を解決し、ユーザのストレスや負担を軽減するために、情報を個人化して提示するシステムを提案、開発している。

問題の原因として、Web サイトが Web ページ単位で構成されていることや、各 Web ページの構成内容や Web ページ間のハイパーリンク構造も固定されていることが挙げられる。

そこで、本研究では、柔軟に Web ページを構成するために、3.1 節に示すコンポーネントと呼ぶ情報の単位で Web サイトを構築し、その Web サイトに適したシステムを実現して問題を解決する。また、システムは、3.2 節に示すユーザの Web サイト内での閲覧情報である視聴履歴を用いて、ユーザの嗜好を得る。そして、ユーザのリクエストに対し、嗜好を反映したコンポーネント集合を利用して Web ページを作成、ユーザに提示する。

3.1 コンポーネント管理モデルの概要

提案モデルは、Web ページを構成する要素としての情報の最小単位であるコンポーネント集合および二つのコンポーネントに関連付けるリレーションの集合で構成する。コンポーネントはコンテンツとコンテナからなる。コンテンツは、ある話題に関する具体的な情報をもつ要素である。コンテナは、複数のコンテンツを管理、まとめる役割を果たす。そして、リレーションは二つのコンポーネント（コンテナとコンテナ、コンテナとコンテンツ、コンテンツとコンテンツ）を意味と重みで関連付ける。意味と重みは、どちらも Web サイト作成者が定義する。

図 1 にコンポーネント管理モデルを用いて構築した Web サイトの例を示す。四角がコンテンツを表し、楕円がコンテナを表す。また、矢印はリレーションを表し、意味と重みに関する情報を伴う。なお、矢印の太さは重みの大きさを表す。

3.2 視聴履歴

視聴履歴とは、ユーザが Web サイト閲覧時に辿ったコンポーネントとリレーションの列である。視聴履歴に記録する情報は、コンポーネントを識別する一意の ID、コンポーネ

ある。Step 6 と Step 7 の間で R_k の評価値が閾値 T_2 を超えなければ、対応するコンポーネントは抽出せず無視するなど、コンポーネント数を制御できることである。その結果、提示する情報量を柔軟に制御できる。また、適切な閾値を用いることで、ユーザに有用な情報のみを提示でき、Web ページの構成も単純化できる。

以下、評価関数の詳細を説明する。評価関数は、ユーザの嗜好と Web ページ作成者の意図を反映するために導入する。評価関数では、次の四点を評価する。

- TOP_C からリレーション R に至るリレーションの意味列と、ユーザの視聴傾向との適合度および視聴履歴における当該視聴傾向の出現数。
- リレーション R の重み。
- ユーザが過去にリレーション R を辿った回数。
- リレーション R が生成されてから経過した時間。

同じ意味をもつリレーションの出現数は、その数が多いほど、ユーザに辿られやすく、リレーションの参照先にあるコンポーネントがユーザに閲覧されやすいことを示す。したがって、出現数が多いほど、そのリレーションを重要なものとし、参照先のコンポーネントを抽出しやすくする。リレーションの重みは、Web サイト作成者が同時に提示したい二つのコンポーネントを関連付ける際に大きくなる。

本システムでは、Web サイト作成者の意図も反映するために、重みの大きいものほど抽出しやすくする。視聴傾向は評価関数の中で活用する。コンポーネント抽出時に、視聴傾向と適合するリレーションの経路が存在する場合、経路上に存在するコンポーネントを同時に提示することで、ユーザの情報閲覧に役立つとみなし、抽出しやすくする。リレーションが生成されてから経過した時間は、情報の新鮮度を考慮するために導入したものであり、経過時間が短いほど、リレーションの参照先にあるコンポーネントを抽出しやすくする。

また、評価関数においては、上記の他に、ユーザがリレーションを過去に辿ったかどうかといった情報や、過去に辿ったのならば、それから経過した時間も考慮して計算可能である。

4.3 コンポーネントの提示手法

本節では、抽出したコンポーネント集合から Web ページを作成する手法について述べる。視聴傾向に基づいて、ユーザにとって適切なコンポーネントを抽出できたとしても、単に結果を提示するだけでは、個人化の効果は小さく、ユーザを満足させることはできない。それは、ユーザにとって、Web ページの情報を容易に把握できないからである。

そこで、本研究では、Web ページの色合いや文字の大きさなど、Web デザインの知識に加え、ユーザが閲覧時に興味をもち、理解しやすいコンポーネントの配置方法についても言及する。

Web ページの作成における指針は次の通りである。1. 閲覧において、ユーザに高度なスキルを要求しないこと。2. ブラウザの種類によらず情報の閲覧が可能なこと。3. ユーザが自分自身を認識可能なこと（ユーザ ID の表示）。4. ユーザが過去に閲覧した情報をすぐに再閲覧できること（履歴の活用）。5. ユーザが閲覧したい情報を容易に見ることができる（リレーションの活用）。6. 情報を単一 Web ページに詰め込みすぎないこと。ただし、6 はコンポーネントの抽出時に、各コンポーネントの情報量を考慮することで、抽出数を制限できる。なお、本研究において、Web ページの構成はユー

ザごとに変化し、また同一ユーザが同じコンポーネントをリクエストしたとしても、リクエスト時のユーザの視聴傾向によって結果の Web ページは変化する。したがって、通常の Web サイトで見られる、ハイパーリンク構造が固定されたサイトマップは存在しない。

効果的な Web ページの作成にあたり、まず、提示する情報を、以下の三つの構成要素に分類する。また、それぞれの情報の詳細も示す。

• Web サイトの基本情報

ユーザが訪問している Web サイトの名称やシンボルとなる画像を提示する。また、ユーザ ID やシステムからログオフするためのハイパーリンク、Web サイトのコピーライトなど、全コンポーネントに共通する事項も提示する。

• Web サイトのナビゲーション情報

ユーザが過去に閲覧したコンテナに関して、そのタイトルを閲覧順序に提示する。また、タイトルからハイパーリンクによって、対象となるコンテナへ移動可能とする。

• 個人化情報

抽出したコンポーネント集合を提示する。基本的には、図 2 に示す通り、コンテナ、コンテナとリレーションで繋がれたコンテンツ集合を一塊とする。そして、それをコンポーネントが存在する限り提示する。なお、抽出しなかったコンポーネントに対するリレーションはコンテンツの内容にハイパーリンクとして埋め込むことができる場合は埋め込み、それ以外はコンテンツの下部にリレーションの意味をハイパーリンクとして提示する。

上記要素からなる抽象化した Web ページを図 3 に示す。これらの要素において、基本情報では静的な情報を提示する。一方、ナビゲーション情報および個人化情報では、ユーザの閲覧時に動的に変化する情報を提示する。以下に、各要素において、上記の指針を反映するレイアウトと装飾において考慮すべきことを示す。

• 基本情報

– レイアウト

もし Web サイトのシンボルとなる画像に Web サイトの名称が示されているならば、文字列としての名称は提示しない。また、ユーザ ID やログオフ情報は、常にユーザの目に付く場所に配置する。

– 装飾

背景色やフォントの色は、白の背景に黒文字など、ユーザが閲覧しやすく派手でない、標準的なものとする。また、背景画像を利用する場合も同様に、派手でなく、文字が読み易いものとする。フォントの大きさと太さは、Web サイトの名称に関しては大きく太くし、その他に関してはブラウザ標準のものとする。

• ナビゲーション情報

– レイアウト

閲覧順に画面上部から配置する。

– 装飾

フォントの色や大きさ、太さはブラウザ標準のものとする。

• 個人化情報

– レイアウト

コンポーネントを上部から配置するだけである。しかし、後述するルールに基づき、レイアウトを変化させることで、単調なレイアウトとなることを回避する。コンポーネント間のリレーションはハイパーリンクとして、重みの大きい順にコンポーネントの右部に提示する。

ー 装飾

フォントの色や大きさ、太さは基本的にブラウザ標準のものとする。ただし、ユーザの情報理解を助けるため、トピックは色や大きさ、太さを少し強調し、タイトルはより強調したものとする。さらに、Web ページの中で最も重要なタイトルを最も強調して提示する。また、情報を囲む境界やテーブルの枠に関しては、太さを標準とし、色を協調する。なお、レイアウト同様、装飾もルールに基づき柔軟に変更可能とする。

個人化情報の柔軟かつ変化に富んだ情報提示を実現するために、本研究ではルールに基づくレイアウト、装飾の指定を可能とする。ルールは、抽出したコンポーネント集合およびそれらに関係付けるリレーション集合、さらに Web サイトそのものに対する評価を活用した条件と、それに対する結果で成り立つ。具体的には、コンポーネントやリレーションの数、コンポーネントに対するユーザの閲覧回数、リレーションの意味や重みを考慮する。

例えば、特定の意味をもつリレーションで結ばれた二つのコンポーネントの配置などを指定できる。

以下に、本研究で定義したルールの条件を示す。なお、ルール適用時の優先度は数字が小さいほど高い。

優先度 1 ある値が与えられたとき、コンポーネント数と値の大小を比較する。

コンテナ数やコンテンツ数、リレーション数との比較も可能とする。

優先度 2 ある値が与えられたとき、コンポーネントの閲覧回数と値を比較する。

優先度 3 ある値が与えられたとき、リレーションの重みと値を比較する。

優先度 4 ある値が与えられたとき、あるコンポーネントに関するリレーションの重みの平均、または総和と値を比較する。

優先度 5 ある文字列が与えられたとき、コンテナのジャンルと文字列の等価性を調べる。

コンテンツのトピックやリレーションの意味に対する調査も可能とする。

優先度 6 二つの文字列が与えられたとき、任意の二つのコンポーネントのジャンル、またはトピックが二つの文字列と一つずつ等価であり、コンポーネント同士がリレーションに関連付けられているかを調べる。

コンポーネントとリレーション間の調査も可能とする。

これらのルールに合致するコンポーネントとリレーションのレイアウトと装飾に関して、それぞれ次の変更を可能とする。

● レイアウト

文章の長さ、画像の大きさ、コンテンツ集合のテーブルによる配置、画像に対するテキストの回り込み、コンテンツの横に隣接した配置、ハイパーリンク（リレーシ

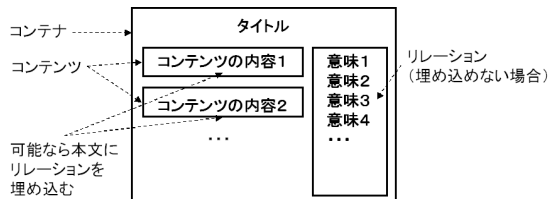


図 2 個人化情報の基本構成

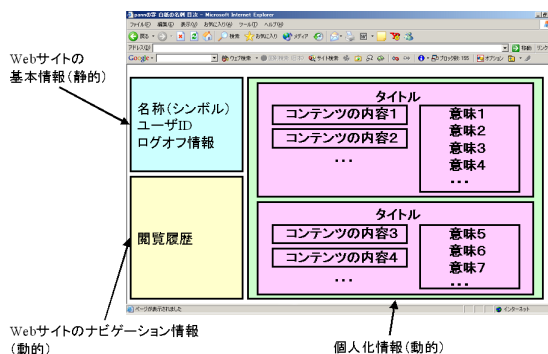


図 3 三要素からなる Web ページの概要

ン)の表示数

● 装飾

フォントの色、フォントの大きさ、フォントの太さ、(ボーダーやテーブルの) 枠線の色、枠線の太さ、枠線の種類

5. 個人化システムの実現

本章では、本研究で実現する個人化システムを簡単に紹介し、その後、前章で述べたコンポーネントの抽出と提示手法の実現について示す。

5.1 個人化システムの概要

提案システムをサーバ、クライアント型のシステムとして実現する。サーバは、ユーザ情報と Web サイトを管理し、クライアント（ユーザ）からのリクエストに応答する。ユーザは、一般的に利用される Web ブラウザを用いて閲覧作業を行うだけである。このことは、ユーザの負担を軽減し、システムを広く利用してもらうために重要である。

サーバで管理するユーザ情報とは ID とパスワード、視聴履歴である。これらは個々のユーザごとに独立しており、したがって、システムの利用においては、ユーザを識別するために、ログイン作業が必要となる。サーバ上で Web サイトを実現するに当たり、コンテナやコンテンツ、リレーションのデータ形式として、XML (Extensible Markup Language)¹²⁾ を利用した。コンテナとコンテンツの定義には RDF (Resource Description Framework)¹³⁾、リレーションの定義には XLink (XML Linking Language)¹⁴⁾ を利用した。

提案システムは、OS が Windows 2000、CPU が Pentium4 2.0GHz、Memory を 512MB 搭載したパソコン上で実装し、動作させた。実装においては、プログラミング言語として Java 2 Platform, Enterprise Edition (J2EE)⁹⁾ を使い、Web サーバアプリケーションとして Tomcat¹⁾ を用いた。また、Web サイトは筆者らの所属する研究室に関して構築し、データの処理においては XML 文書用の問い合わせ言語である XQuery の実装として Quip⁸⁾ を用いた。抽出したコンポーネント集合を利用した Web ページの作成には XML 文書用のスタイルシート言語である XSLT¹⁵⁾ と、HTML 文書の見栄えを定

義するスタイルシート言語である CSS¹¹⁾ を用いた。なお、構築した Web サイトは、25 のコンテナ、104 のコンテンツ、278 のリレーションからなる。

5.2 コンポーネント抽出部の実現

ユーザからのリクエストが発生したとき、システムでは視聴傾向に基づいて、4.2 節に示す手順で、ユーザに適したコンポーネント集合を抽出する。本研究では今回、コンポーネントの抽出数の上限を 5 とした。また、視聴傾向はリクエスト毎に更新した。コンポーネントの抽出においては、評価式 $E(TOP_C, R)$ が重要である。本研究では、評価式 $E(TOP_C, R)$ を式 1 の通り実装した。

$$E(TOP_C, R) = \alpha \min \left(1, \frac{p(TOP_C, R) - THRESHOLD_p}{2THRESHOLD_p} \right) + \beta \frac{w}{MAX_w} + \gamma \frac{\max((MAX_a - a), 0)}{MAX_a} \quad (1)$$

なお、式 1 における関数や変数、定数の意味は次の通りである。

- $p(TOP_C, R)$, $THRESHOLD_p$: 関数 p は、Web サイトにおいて、ユーザにリクエストされたコンポーネントからリレーション R までの経路となるリレーション集合の意味列を考えたとき、その意味列が視聴傾向として存在するかを判別する。存在するなら、視聴履歴中での出現回数を返す。存在しない場合は 0 を返す。定数 $THRESHOLD_p$ は視聴傾向を抽出する際に利用する閾値でもある。本研究では、視聴履歴中に同じ意味列が閾値以上に出現するときに、その意味列を視聴傾向とみなす。
- w , MAX_w : 変数 w はリレーション R の重みである。定数 MAX_w はユーザが定義できる重みの最大値である。
- a , MAX_a : 変数 a はユーザが過去の閲覧においてリレーション R を辿った回数である。定数 MAX_a は、情報の理解に必要な十分だとみなす閲覧回数の最大値である。
- α , β , γ : 評価式において、各項の影響力を決定する係数である。 $\alpha + \beta + \gamma = 1$ とする。

式 1 において、第一項は視聴傾向の影響、第二項はリレーションの重みの影響、第三項はコンポーネントの閲覧回数の影響をそれぞれ考慮した。なお、コンポーネント作成時からの経過時間の影響は無視した。また、各値はそれぞれ $THRESHOLD_p = 2$, $MAX_w = 3$, $MAX_a = 5$ とした。 $THRESHOLD_p = 2$ であるため、視聴傾向を最も抽出しやすい設定である。係数はそれぞれ $\alpha = 0.4$, $\beta = 0.4$, $\gamma = 0.2$ とし、視聴傾向と重みを重視した。つまり、ユーザと Web サイト作成者それぞれの嗜好や意図を反映した抽出となる。

5.3 個人化した Web ページの作成

本節では、前章の 4.3 節に示した、個人化した Web ページの作成に関する実装について述べる。

まず、図 2 の構造にしたがい、各コンテナに対して、抽出結果でコンテナの子孫となるコンテンツを振り分ける。また、振り分けと同時に、提示する Web ページおよび直前に提示した Web ページに含まれるコンポーネントが参照先となるリレーション以外をハイパーリンクになりえるリレーションとする。次に、Web サイト作成者の定義したルールを適用し、一つの XML 文書を作成する。最後に、XML 文書を XSLT

を用いて HTML 文書に変換し、CSS を用いて装飾を行う。

提示において、想定するブラウザの縦幅と横幅はそれぞれ 600 ピクセル、800 ピクセルとする。そこに、前章で示した Web ページにおける三つの構成要素の配置位置を適用する。Web サイトの基本情報はブラウザの画面左上に、縦幅と横幅をそれぞれ 300 ピクセル、200 ピクセルに固定して配置する。ナビゲーション情報は画面左下に、横幅を 200 ピクセルに固定して配置する。個人化情報は画面の中央部に縦幅、横幅を固定せずに配置する。

構成要素の配置方法として、Web ページをフレームによって区切り、位置を指定する方法と、CSS を用いて配置を指定する方法がある。本研究では、配置に対する柔軟性を考慮して、CSS によって配置を実現した。以下に Web ページにおける各種設定を示す。これらの設定も CSS によって実現しており、括弧で囲む文字列は、CSS における色や文字の太さの定義を示す。

• フォント

通常の文字色は黒色 (#000000) とする。強調色は赤色 (#ff0000) とし、ハイパーリンクに関する文字色はブラウザ標準のものとする。また、トピックは栗色 (#800000)、タイトルは濃紺色 (#000080) とする。文字の大きさに関しては、ブラウザ標準のものを基準とし、タイトルを 120% の大きさ、さらにタイトルの中で、Web ページにとって最も重要なものは 150% の大きさとする。太さは、強調する場合とトピック、タイトルに関して太字 (bold) とする。

• 背景

背景色は白色 (#ffffff) とする。

• 境界とテーブルの枠

境界の太さはブラウザ標準 (medium) とし、線の形状は点線 (dotted)、色は緑色 (#008000) とする。テーブルの枠は表示しない。

Web ページの作成において、実際に定義したルールと、ルールを適用した結果について述べる。なお、ルールを適用する対象は抽出したコンポーネント集合である。ルールによっては、Web サイト全体が適用対象となる場合もある。

• ルール 1

内容 ユーザが一度以上コンポーネントを閲覧したことがある。

結果 コンポーネントのフォントの大きさをブラウザ標準のものに対して 90% とする。

• ルール 2

内容 リレーションの重みが 2 より大きい。

結果 リレーションの参照先のコンポーネントのフォントの大きさをブラウザの標準に対して 110% とする。

• ルール 3

内容 トピックが「画像」であるコンテンツが存在する。

結果 列数を 3 に設定したテーブルに画像コンテンツを配置する。

• ルール 4

内容 トピックが「画像」と「詳細」である二つのコンテンツが存在し、リレーションで直接結ばれている。

結果 画像を左、詳細のテキストを右から回り込みで配置する。

図 4 と図 5 に、システムによって個人化、作成した Web

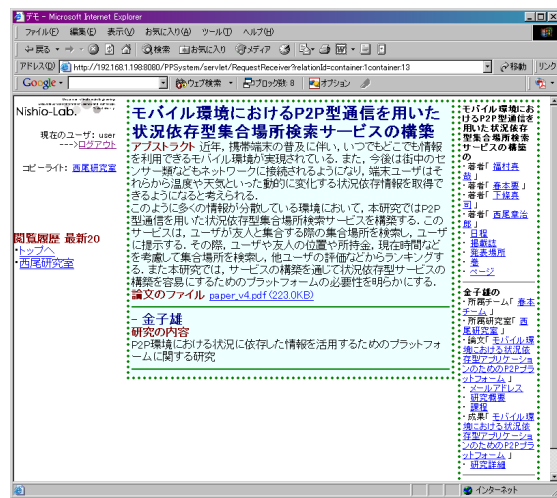


図 4 システムが作成した Web ページ 1

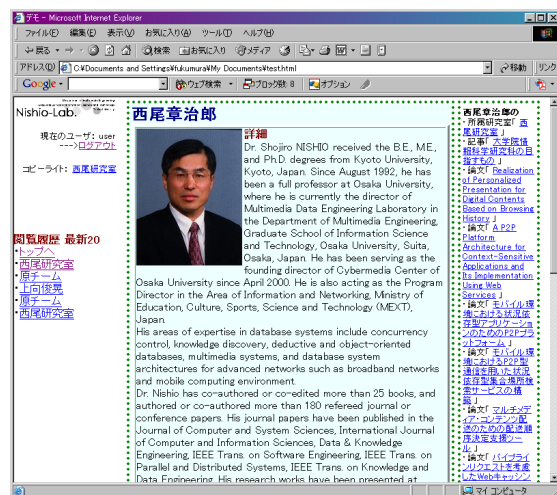


図 5 システムが作成した Web ページ 2

ページを示す。図 4 は、論文とそのアブストラクトや著者を提示し、図 5 は研究室の構成員の一人に関する説明と画像を提示している。図 5 においては、ルール 4 に従い、画像に回り込む形で説明を提示している。また、両画像において、画面左部には、Web サイトとユーザに関する情報およびユーザの閲覧順序を提示し、画面右側には、各コンポーネントに関する関係を、ユーザが理解しやすい形で提示している。次章では、これらの個人化された Web ページおよびシステムに対する評価について述べる。

5.4 システム評価案

本研究では、ユーザが Web サイトにおいて、興味のある情報を容易に閲覧でき、内容も容易に理解できるシステムの構築を目的としている。また、Web サイト作成者が Web サイトに対して、作成の意図を反映できるシステムの構築も目的である。したがって、本システムの評価は、ユーザと Web サイト作成者それぞれの視点から行う必要がある。

まず、ユーザ側からの評価方法を考える。本システムはあらかじめ用意してある Web サイトを個人化する手法とは異なるため、従来の個人化システムとの比較評価は不適当である。そこで、本研究ではまず、本システムに対して、ユーザの側から二つの方法での評価を行う。

一つ目は、本システムにおいて、ユーザの視聴傾向を考慮する場合と考慮しない場合の閲覧において、Web サイト内

表 1 アンケート項目

Web ページに対する操作の分かりやすさ
履歴に対する操作の分かりやすさ
Web ページの構成の分かりやすさ
Web ページの見易さ (色, フォント, ハイパーリンク数, 情報量など)
提示された情報のまとまりの良さ
リクエスト時の反応の良さ
システムに対する全体的な好感度
システムの全体的な役立ち度

に存在する、決められたコンポーネントを閲覧してもらう方法である。その際、視聴傾向はあらかじめ設定しておく。これは、Web ページの個人化に視聴傾向を反映することで情報の見つけやすさが変化するかの評価であり、それぞれの場合でコンポーネントを閲覧するまでにかかった作業量を比較する。また、情報の見つけやすさに関してアンケートをとり、ユーザの主観で評価してもらう。

二つ目は、一つ目と同様、視聴傾向を考慮する場合と考慮しない場合において、ユーザに Web サイト内のコンポーネントを自由に閲覧してもらう方法である。これは、システムがユーザの閲覧作業に伴う視聴傾向の抽出と Web ページの個人化を効果的に実現しているかの評価である。こちらは、閲覧のしやすさと情報の理解しやすさに関してアンケートを行い、主観で評価してもらう。さらに、表 1 に示す項目に関してもアンケートを行う。

Web サイト作成者側からの評価には、Web サイト作成者が、ユーザに見せたいコンポーネントをユーザに効果的に提示できているかを考慮する。そこで、コンポーネントの抽出ごとに、考えられるリレーションの重みの最大値の和を分母、実際に利用したリレーションの重みの和を分子とした式により値を導出する。この値が 1 に近いほど高評価とする。システムが作成した各 Web ページに対して、高評価となる Web ページが多いとき、Web サイト作成者の意図が十分に反映されているとみなす。

なお、Web サイトの作成において、作成時間など、ユーザの手間も評価する必要がある。

6. 今後の研究

今回、本システムはサーバ、クライアント環境を想定し、Web ページ作成時に利用するスタイルシートは PC を対象としている。システムを PDA をはじめとするモバイル端末に対応させる利点として、少ない操作および小さな画面で、ユーザが求める情報を簡単に見つけられることが挙げられる。また、コンポーネントから Web ページを作成するため、例えば、動画を再生できない端末に対しては、Web ページの構成要素の抽出時に動画を含めないといった変更が容易に可能である。

モバイル端末への対応には、各端末用のスタイルシートを用意する必要がある。この作業は Web サイト作成者の負担となるが、通常の Web サイトを作成する際にも同様の作業は必要である。さらに、システムとして、端末の種類の判別し、適当なスタイルシートを適用する必要もある。

将来的には、本システムのモバイル P2P 環境における適用も考えている。ユーザの状況に応じて、ピアに蓄えられているコンポーネントの中から、必要なものを取得し、個人化するサービスを適当なピアが提供するものとする。情報は分散しており、静的なリレーションによって関連付けたコン

ポーネント集合と、動的なリレーションで関連付けられるコンポーネント集合を考える。

動的なリレーションで関連付ける理由は、特にモバイル P2P 環境での不安定なネットワーク状況に対応するためである。また、ピア数が増加すると、それに伴いコンポーネント数も増大し、リレーション情報を一括して管理することが困難となるからである。

動的なリレーションとは、サービスに対してリクエストが発生したとき、抽出可能なコンポーネント間をルールによって動的に関連付けるものである。動的な処理を行うことで、リアルタイムでユーザーに適した情報を提供できる。

抽出においては、応答時間やコンポーネントの品質、ピア間の回線状況を考慮して結果に反映するなど、P2P 環境に対する新たな処理の仕組みも必要となる。

7. おわりに

本研究では、ユーザーの嗜好に合わせ、Web ページを動的に生成し提示するシステムを実現した。まず、個人化を考慮した Web サイトモデルとして、Web ページ単位ではなく、コンテンツやコンテンツ、リレーションといった単位からなるコンポーネント格納モデルを提案した。次に、ユーザーの視聴傾向に基づく、ユーザーに適したコンポーネントの抽出手法と、個人化した Web ページの動的な作成手法を提案した。

本システムを用いることで、ユーザーに役立つ情報のみで Web ページを作成したり、ユーザーが欲する情報に辿り着くまでの作業量を削減できる。また、Web ページの提示においても、ルールに基づくコンポーネントの配置を可能とし、Web ページが単調にならず、ユーザーが興味をもって Web ページを閲覧できるように工夫した。

今後は、まず、本システムを利用した評価実験を行う。その後、既存のツールでは提案モデルに基づくデータの作成は困難であるため、情報提供者を支援するオーサリングツールを実現する予定である。また、視聴履歴においてリレーションの意味の全く同じ辿り方のみではなく、ノイズを含む辿り方でも視聴傾向を抽出できる手法を考える。さらに、6章で述べた、PDA などのモバイル端末への対応や、モバイル P2P 環境におけるシステムの実現も考えている。

謝 辞

本研究は、日本学術振興会科学研究費補助金基盤研究 (A)「デジタルコンテンツとしての懷徳堂研究」(課題番号 13309011)と文部科学省科学技術振興調整費「モバイル環境向 P2P 型情報共有基盤の確立」によっている。また、本研究の一部は、文部科学省 21 世紀 COE プログラム「ネットワーク共生環境を築く情報技術の創出」(研究拠点形成費補助金)の研究助成によるものである。ここに記して謝意を表す。

参 考 文 献

- 1) Apache Software Foundation: “The Jakarta Site – Apache Tomcat –,” <http://jakarta.apache.org/tomcat/>.
- 2) CAST: “CAST: Center for Applied Special Technology,” <http://www.cast.org/>.
- 3) David Durand and Paul Kahn: “MAPA: A System for Inducing and Visualizing Hierarchy in Websites,” Proc. of the International Conference on HyperText 1998, pp. 66–76, Pittsburgh (June 1998).
- 4) Ed H. Chi, Peter Pirolli, Kim Chen, and James

- Pitkow: “Using Information Scent to Model User Information Needs and Actions on the Web,” Proc. of ACM CHI 2001 Conference on Human Factors in Computing Systems, pp. 490–497, Seattle (April 2001).
- 5) HYPERGENIC: “Hypergenic WebNize 1000 Metro,” <http://metro.asb.dk/webnize/metro/>.
- 6) Motoki Miura, Buntarou Shizuki, and Jiro Tanaka: “inlineLink: Inline Expansion Link Methods in Hypertext Browsing,” Proc. of International Conference on Internet Computing (IC2001), Vol. II, pp. 653–659 (June 2001).
- 7) NTT Cyber Solutions Laboratories: “YoriDoriWeb,” <http://www.cyber-trial.com/yoridori/hp/>.
- 8) Software AG: “Tamino Developer Community – W3C XQuery Information –,” <http://developer.softwreag.com/tamino/quip/>.
- 9) Sun Microsystems: “The Source for Java Technology,” <http://java.sun.com/>.
- 10) Tingshao Zhu, Russ Greiner and Gerald Häubl: “An Effective Complete-Web Recommender System,” Proc. of the 20th International World Wide Web Conference, Budapest (May 2003).
- 11) World Wide Web Consortium: “Cascading Style Sheets,” <http://www.w3.org/Style/CSS/>.
- 12) World Wide Web Consortium: “Extensible Markup Language (XML) 1.0 (Second Edition),” <http://www.w3c.org/TR/REC-xml/>.
- 13) World Wide Web Consortium: “Resource Description Framework (RDF) Model and Syntax Specification,” <http://www.w3c.org/TR/REC-rdf-syntax/>.
- 14) WORLD WIDE WEB consortium (W3C): “XML Linking Language (XLink) Version 1.0,” <http://www.w3c.org/TR/xlink/>.
- 15) WORLD WIDE WEB consortium (W3C): “XSL Transformations (XSLT) Version 1.0,” <http://www.w3c.org/TR/xslt/>.
- 16) Yahoo Japan Corporation: “My Yahoo!,” <http://my.yahoo.co.jp/>.
- 17) 清光英成, 竹内淳記, 田中克己: “アクティブルールを用いた Web ページの動的再構成,” Proc. 情報処理学会研究報告, Vol.100, No.69 2000–DBS–122–50, pp. 383–390 (July 2000).
- 18) 竹内淳記, 清光英成, 田中克己: “アクセス履歴の集約とメタルールに基づく Web コンテンツのアクセス管理機構,” Proc. 情報処理学会研究報告, Vol.100, No.69 2000–DBS–122–41, pp. 315–322 (July 2000).
- 19) 福村真哉, 中野賢, 春本要, 下條真司, 西尾章治郎: “視聴履歴に基づく個人化された提示手法,” 電子情報通信学会データ工学ワークショップ (DEWS 2002) 論文集, (Mar. 2002).