

視聴履歴に基づくデジタルコンテンツの個人化された提示手法

福村 真哉[†] 中野 賢[†] 春本 要[‡] 下條 真司[‡] 西尾 章治郎[†]

[†] 大阪大学大学院工学研究科 情報システム工学専攻

[‡] 大阪大学 サイバーメディアセンター

E-mail: {fukumura,tnakano,nishio}@ise.eng.osaka-u.ac.jp

{harumoto,shimojo}@cmc.osaka-u.ac.jp

概要

近年、テキストや画像、音声など、様々なデータがデジタル化され、インターネット上で広く公開されている。特に、WWW (World Wide Web) を用いた情報の提供は、現在、多くの人に利用されている。しかし、現在の Web サイトは、各 Web ページの内容や Web ページ間のリンク構造が固定的であり、ページ閲覧者の閲覧目的や嗜好に必ずしも合致しない構成となっている場合がほとんどである。したがって、ページ閲覧者は、興味のある情報に辿り着くまでに、何度もリンクを辿ったり、不必要な情報を閲覧する必要が出てくる。また、現在の Web サイトでは、ページ閲覧者の多様な閲覧目的や嗜好に対し、適切な情報を提示することは困難である。そこで、本研究では、ページ閲覧者の興味に合わせ、Web ページを動的に生成し、提示するシステムの実現を目的とする。まず、ページ閲覧者に適した情報で、Web ページを生成するために、ページ単位ではなく、情報の素材(コンテンツ)単位で構成する新たなデータ格納モデルを提案する。次に、ページ閲覧者の興味を取得するために、素材間の関係における意味や強さを考慮した、視聴履歴の利用方法を提案する。

1 はじめに

近年、美術館や博物館に展示されている作品を保存したり、作品や企業の製品に関する情報をインターネット上で多くの人へ公開するために、テキストや画像、音声など、あらゆる情報のデジタル化が進んでいる。また、インターネット上における情報の公開には、情報発信の基盤として大いに注目され、急速に普及している WWW (World Wide Web) において、Web サイトを構築する方法が一般的となっている。そして、公開されている情報に対し、パソコンや携帯電話、PDA (Personal Digital Assistants) など、Web ブラウジング機能を備えた様々な端末を用いて、誰でも容易にあらゆる場所から閲覧できる環境が整いつつある。そのため、現在、WWW における情報提供は、多くの人にとって、生活に欠かせないものとなっている。

しかし、現在の Web サイトは、各 Web ページの内容や Web ページ間のリンク構造が固定的であり、ページ閲覧者の閲覧目的や嗜好に必ずしも一致しない構成となっている場合がほとんどである。したがって、ページ閲覧者は、興味のある情報に辿り着くまでに、何度もリンクを辿ったり、不必要な情報を閲覧する必要があり、ストレスや負担を感じる。また、現在の Web サイトでは、ペー

ジ閲覧者の多様な閲覧目的や嗜好に対し、適切な情報を提示することは困難である。

そこで、本研究では、ページ閲覧者の興味に合った Web ページを動的に生成し、提示することで、ページ閲覧者が効率的に情報を取得できるシステムの実現を目的とする。Web ページの生成においては、特に、情報の提示方法とページ構成を個人化する。本研究では、まず、ページ閲覧者に適した情報だけで Web ページを生成するために、ページ単位ではなく、情報の素材単位で構成する新たなデータ格納モデルを提案する。また、ページ閲覧者の興味を取得するために、素材間の関係における意味や強さを考慮した、視聴履歴の利用方法も提案する。

以下、第2章で既存 Web サイトの問題点について述べ、新たなデータ格納モデルを提案する。その後、第3章で Web ページの動的生成方法を提案し、第4章で提案モデルの実現方法を示す。そして、第5章で関連研究を紹介し、第6章で本研究についてまとめる。

2 個人化を考慮したデータ格納モデル

本章では、まず、WWW の特徴と一般的な Web サイトモデルについて述べ、Web ページの閲覧における問題点を挙げる。次に、問題を解決するための指針を示し、そ

表 1: 視聴履歴の構成要素

要素	意味
cr(id)	ID が id であるコンテナが参照されたことを示す。
ct(id)	ID が id であるコンテンツが参照されたことを示す。
r(meaning, priority)	意味に対応する数字が meaning である意味と、強さが priority であるリレーションが辿られたことを示す。
a(jump)	過去に閲覧したコンテナやコンテンツへの移動を示す。

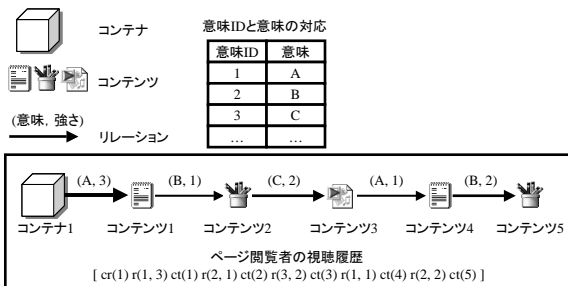


図 3: 視聴履歴の例

なパターンを視聴傾向として検出する。なお、視聴傾向の検出時期としては、一定期間ごと、視聴履歴が一定量蓄積されるごと、ページ閲覧者の訪問回数が一定数に達するごとなどが考えられるが、検出作業を定期的に行うことで、ページ閲覧者の嗜好をより反映した Web ページが生成できる。

視聴傾向の検出においては、次に示す点に着目する。

- 視聴傾向の出現回数
視聴履歴において、視聴傾向の出現回数が多いものほど、ページ閲覧者によって辿られやすく、重要であると考えられる。
- 情報の新鮮度
視聴履歴の古い情報が視聴傾向の検出作業に与える影響を小さくすることで、ページ閲覧者の興味の推移を反映する。

3.2.2 視聴傾向に基づくデータの抽出

コンテナやコンテンツに対して、ページ閲覧者のリクエストが発生した時、前節で述べた視聴傾向に基づき、Web サーバに蓄えられたデータの中から、ページ閲覧者に適したものを抽出する。

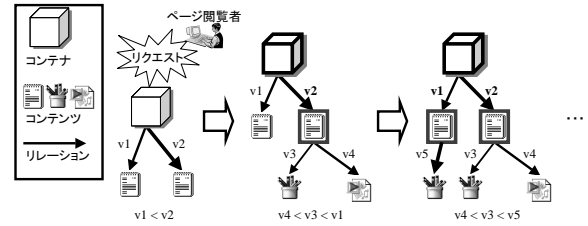


図 4: 視聴傾向に基づくコンテナ，コンテンツの抽出

抽出作業においては、ページ閲覧者の閲覧特性を反映し、かつ、情報提供者の意向も反映した Web ページを生成するために、「リレーションの強さ P 」や「リクエスト位置からのリレーションのパス数 D 」、「同じ意味のリレーションを辿った回数 N 」、「リレーションの辿り方及び辿った回数 S 」に着目する。

抽出作業では、まず、リクエストされたコンテナまたはコンテンツを参照元とする各リレーションに対して、上記の要素を利用した「評価値 v 」を求め、値の最も大きいリレーションの参照先を抽出結果に加える。その後、結果に加えたコンテナやコンテンツを参照元とするリレーションがあれば、それらを評価対象に加えて評価値を求め、再び値の最も大きいリレーションの参照先を抽出結果に加える。そして、以上の作業を評価対象となるリレーションがなくなるか、十分なデータを抽出するまで繰り返す(図 4)。

ここで、ページ閲覧者の閲覧特性を反映するために、式 1 で求める v において、それぞれのパラメタの影響を変化させることを考える。例えば、式 1 において、パラメタ S と N の影響を考えなければ、単純に良く辿るリレーションで結ばれたデータを提示するよりも、全体の流れを重視してデータを提示することを意味する。

$$v = f(P, D, N, S) \quad (1)$$

3.2.3 ページレイアウトの決定

視聴傾向に基づいて、コンテナやコンテンツを抽出できても、単純に結果をまとめて提示するだけでは、ページ閲覧者を満足させることは難しい。そこで、抽出結果をまとめる際に、Web デザインの知識を踏まえたページレイアウトを活用して、ページ閲覧者が興味を持って閲覧できる Web ページを実現する。その際、例えば、リレーションの強さを考慮して、関連性の深い情報を近くに配置すれば、まとまりがあり、閲覧しやすい Web ページを提示できる。

4 データ格納モデルの実現

本章では、第 2 章で提案したデータ格納モデルに基づき、コンテナやコンテンツ，リレーションを実現する方

法について考察する。

通常、一般的な Web サイトは、HTML で記述され、ハイパーリンクで結ばれた Web ページ群で構築されている。しかし、これらの技術だけでは、コンテナやコンテンツ、リレーションを実現することは難しい。そこで、本章では、一つの実現方法として、コンテナやコンテンツ、リレーションを XML (Extensible Markup Language)[2] に基づく技術で定義することを考える。

XML に基づく技術を利用するメリットとして次のことが挙げられる。

- XML はデータ構造だけを定義するものであり、その表示方法までは関与しない。したがって、コンテナやコンテンツを XML で定義することで、Web ページを生成する際に、構造とレイアウトを全く独立して定義できる。
- XML は HTML とは異なり、タグの名称やデータ構造を自由に定義できる。したがって、複数の情報提供者で作業する場合など、提供する情報に適したタグを用意し、可読性の高くすることで、作業を効率化できる。
- メタ情報表現のフレームワークである RDF (Resource Description Framework)[7] を用いることで、提供する情報そのものとメタデータを切り離すことができる。その結果、提供する情報とメタデータの両方を同時に管理する必要がなくなり、情報提供者の管理作業における負担を軽減できる。
- WWW 上の 2 つ以上のリソースを関連付けるための規格である XLink (XML Linking Language)[8] を用いる事で、コンテナやコンテンツの関係において、ハイパーリンクでは定義できない、複雑な定義が可能となる。

一方、デメリットとして、現在利用されている HTML ファイルがそのまま流用できないことが挙げられる。

4.1 コンテナ、コンテンツの定義

コンテナとコンテンツは、RDF で定義したメタ情報と、情報提供者が自由に定義できる本体の二つの要素で構成する。

コンテナやコンテンツのメタ情報は、ID やジャンル、トピックなど、本論文で独自に定義したスキーマを利用して定義する。また、本体に関しては、特にタグの名称や構造に指定はなく、XML の構文さえ正しければ、記述する内容は情報提供者が自由に定義できる。さらに、コンテナは、コンテンツをまとめる役割を果たす性質上、本体に情報をもたず、XML 宣言を行っただけのファイルとなる。なお、コンテンツの本体では、テキスト情報の他に画像ファイルや音声ファイルなどの外部ファイルも指

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<RDF:RDF xmlns:RDF=
"http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
xmlns:kms="http://www.kaitokudo.jp/meta-scheme/">
<RDF:Description
RDF:about="http://.../container1.xml">
<kms:container kms:id="1">
<kms:genre>人物</kms:genre>
<kms:signage>
<RDF:Seq>
<RDF:li>三宅石庵</RDF:li>
<RDF:li>みやげせきあん</RDF:li>
</RDF:Seq>
</kms:signage>
<kms:dateInformation>
<kms:creation kms:date="2002-02-13"/>
<kms:renewal kms:date="2002-02-13"/>
</kms:dateInformation>
</kms:container>
</RDF:Description>
</RDF:RDF>
```

図 5: コンテナのメタ情報の定義例

定できるが、そのときは外部ファイルもまとめて本体とみなす(図 2)。

図 5 に、あるコンテナのメタ情報の定義例を示す。図 5 のタグや属性において、“RDF” を含むものは、RDF で定義されているタグや属性であり、“kms” を含むものは、本研究で独自に定義したタグや属性である。

メタ情報では、RDF の構文に従い、“RDF:Description” の属性 “RDF:about” で、コンテナの本体が存在する URI (Uniform Resource Identifiers) を指定する。そして、その子要素で、コンテナとコンテンツの識別やジャンルやコンテナの表記、作成日、更新日を記述する。

図 5 に、あるコンテンツのメタ情報の定義例を示す。コンテンツのメタ情報は、基本的にコンテナと同じであるが、“kms:topic” でトピックを記述し、さらに、“kms:keyword” タグを用いて、そのコンテンツに関するキーワードを複数記述できる点で異なる。

4.2 リレーションの定義

リレーションは、リンクの種類や参照先、参照元などを指定でき、ハイパーリンクよりも柔軟な定義が行える XLink で定義する。図 7 にリレーションの定義例を示す。

図 7 では、まず、“kms:locatorDefinition” において、その子要素である “kms:locator” を用いて、コンテナやコンテンツ本体の URI に一意に識別可能なラベル付けを行う。ロケータ定義を行うことで、もし、本体の URI が変更されたとしても、ロケータ定義部を編集するだけで良く、変更作業の簡単化、情報提供者の負担の軽減が期待できる。次に、リレーション定義部である “km-

```

<?xml version="1.0" encoding="Shift_JIS" ?>
<RDF:RDF xmlns:RDF=
"http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
xmlns:kms="http://www.kaitokudo.jp/meta-scheme/">
<RDF:Description RDF:about="http://.../content1.xml">
<kms:content kms:id="1">
<kms:topic>画像</kms:topic>
<kms:dateInformation>
<kms:creation kms:date="2002-02-13"/>
<kms:renewal kms:date="2002-02-13"/>
</kms:dateInformation>
<kms:keyword>
<RDF:Bag>
<RDF:li>キーワード 1</RDF:li>
<RDF:li>キーワード 2</RDF:li>
</RDF:Bag>
</kms:keyword>
</kms:content>
</RDF:Description>
</RDF:RDF>

```

図 6: コンテンツのメタ情報の定義例

s:relationDefinition”において、コンテナやコンテンツ間の関係を、“kms:relation”で定義する。このタグは、参照元と参照先、意味と強さを情報としてもつ。

なお、提案モデルにおいて、リレーションをコンテナやコンテンツの定義とは独立に定義しているが、これは、複数人の情報提供者で作業するとき、各情報提供者がコンテナやコンテンツ同士の繋がりを的確に把握でき、管理が容易となるからである。また、意味や強さは、基本的に各 Web サイトごとに、情報提供者が定義するが、美術館や博物館など、意味を共有できる場合は、意味を標準化することが望ましい。

4.3 プロトタイプシステムの動作例

本研究では、大阪大学の源流と言われる「懐徳堂」に関する人物や器物、書物などの貴重資料に関するデジタルコンテンツ群を利用し、提案モデルに基づいたコンテナやコンテンツ、リレーションを作成した。データ数はそれぞれ 45,281,409 であった。ここで、リレーションの意味としては、表 2 を用い、リレーションの強さは 1~3 の三段階で定めた。また、提案した Web ページの動的生成手順に従って Web ページを生成するプロトタイプシステムを実現した。なお、プロトタイプシステムでは、視聴傾向の検出と視聴傾向に基づくデータの抽出、レイアウトの決定を以下の方法で行った。

- 視聴傾向の検出

視聴履歴において、リレーションの意味の並びのみを考え、同じ並びが 2 つ以上見つかった場合に、その並びを視聴傾向として検出した。

```

<?xml version="1.0" encoding="Shift_JIS" ?>
<kms:relationInformation xmlns:xlink=
"http://www.w3c.org/1999/xlink">
xmlns:kms="http://www.kaitokudo.jp/meta-scheme/">
<kms:dateInformation>
<kms:creation kms:date="2001-12-21"/>
<kms:renewal date="2001-12-21"/>
</kms:dateInformation>
<kms:locatorDefinition xlink:type="extended"
xlink:title="ロケータ定義">
<kms:locator xlink:type="locator"
xlink:href="http://.../container1.xml"
xlink:label="container:1"/>
<kms:locator xlink:type="locator"
xlink:href="http://.../container2.xml"
xlink:label="container:2"/>
<kms:locator xlink:type="locator"
xlink:href="http://.../content1.xml"
xlink:label="content:1"/>
<kms:locator xlink:type="locator"
xlink:href="http://.../content2.xml"
xlink:label="content:2"/>
</kms:locator>
<kms:relationDefinition xlink:type="extended"
xlink:title="リレーション定義">
<kms:relation xlink:type="arc"
xlink:from="container:1" xlink:to="container:2"
xlink:title="作成者 (2)"/>
<kms:relation xlink:type="arc"
xlink:from="container:1" xlink:to="content:1"
xlink:title="画像 (3)"/>
<kms:relation xlink:type="arc"
xlink:from="content:1" xlink:to="content:2"
xlink:title="拡大画像 (1)"/>
<kms:relation xlink:type="arc"
xlink:from="content:2" xlink:to="container:2"
xlink:title="器物 (3)"/>
</kms:relation>
</kms:relationInformation>

```

図 7: リレーションの定義例

- 視聴傾向に基づくデータの抽出

抽出における評価値 v は、式 1 において、 P, D, S を考慮して設定した式 2 で求めた。式 2 における p と d は、評価対象となるリレーションの強さとリクエスト位置からのリレーションのパス数を表す。また、 p と d の最大値である Max_p と Max_d は、今回、3 と 6 に設定した。さらに、 s は、検出した視聴傾向を出現回数の少ない順に並べた時の順番であり、 Max_s はその最大値を示す。ここで、今回、視聴傾向が存在しないリレーションに対する s の値は 0 とした。評価においては、特に、リレーションの強さを結果に強く反映する設定とした。なお、最終的に抽出可能なコンテナとコンテンツの最大数は 10 とした。

$$v = \left(1 + \frac{s}{Max_s}\right) \left(0.9 \frac{p}{Max_p} + 0.1 \frac{Max_d - d}{Max_d}\right) \quad (2)$$

表 2: リレーションの意味

詳細	画像	拡大画像	書物
関係人物	研究	年代	解題
用語	解題補助	器物	漢籍分類

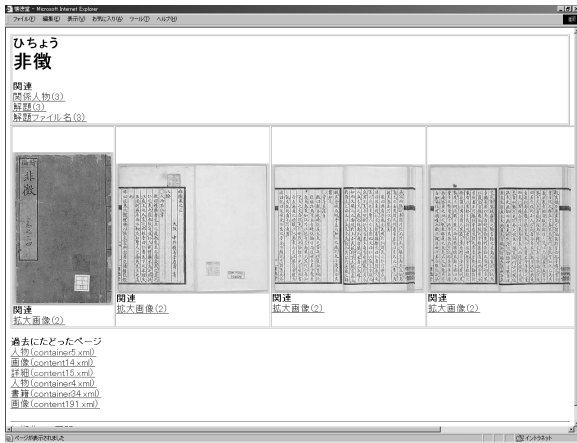


図 8: ページ閲覧者 A に対する Web ページ

● ページレイアウトの決定

今回、Web ページの生成においては、Web デザインやリレーションの強さを考慮せず、ページ閲覧者がアクセスしたコンテナやコンテンツに対して、リレーションのパス数が小さいものをページ上部に配置した。なお、レイアウトの定義は、XML のスタイルシートである XSLT (XSL Transformations)[9] を利用した。

上記の設定において、視聴目的として「書物」の次に「画像」を辿るページ閲覧者 A と「書物」の次に「関係人物」を辿り、その後「詳細」を辿るページ閲覧者 B を想定して、プロトタイプシステムの動作確認を行った。

その結果、プロトタイプシステムは、ページ閲覧者 A に対しては図 8 の Web ページ、ページ閲覧者 B に対しては図 9 の Web ページを生成した。これらの図から、プロトタイプシステムは、ページ閲覧者の視聴傾向に基づいて、ページ閲覧者に適した Web ページを個人化して生成していることが確認できる。

5 関連研究

本章では、Web ページの提示方法やページ閲覧者の行動とページデザイン、Web サイトの効果的な閲覧手法に関する研究を紹介し、本研究との違いを説明する。

Web ページの効果的な閲覧方法として、三浦らはリンク先の文書をリンク元文書に埋め込む手法 [5] を提案している。これは、リンクをクリックしたときにリンクの



図 9: ページ閲覧者 B に対する Web ページ

参照先に移動するのではなく、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が、提示する情報量が多くなる場合、内容の把握が困難となり、必要な情報を見失うおそれがある。

Web ページの効果的な提示方法として、Metro[4] や CAST[1] といったプロジェクトがある。前者は個々の 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、ページ閲覧者に対して、Web サイトの構成を地下鉄の路線図に見立てて提示する。後者は、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。どちらも Web サイトの構成を把握し、目的の情報の取得に役立つが、ページ内容やサイト構造は固定的であり、個人化は行われていない。

ページ閲覧者の行動と Web ページのデザインに関する研究として、Cheese[3] や清光、竹内らの研究 [11, 12] がある。前者は、マウスの動作をクライアント端末でトレースし、その情報をサーバに収集、解析することで、多くのページ閲覧者に共通する傾向を検出して、Web ページのインターフェイスデザインに役立てる研究である。後者は、ページ閲覧者が過去に閲覧した Web ページや、辿ったリンクなどの情報に基づき、Web ページに提示する情報やリンク構造をルールを用いて動的に変化させる方法について議論している。前者は Web ページの辿り方について言及しておらず、後者は Web ページの構成を動的に変更できるものの、あらかじめオリジナルの Web ページを用意しておく必要がある点で、本研究では異なる。

個人化した Web ページを動的に生成するサービスとして、IT Pro の My Page[6] や My Yahoo![10] がある。これらのサービスは、ページ閲覧者の設定した趣味やページレイアウトなどに従って、Web ページを提供する。しかし、様々なカスタマイズが行える反面、設定作業がペー

ジ閲覧者の負担となる。また、Web ページの生成においては、情報間の関係や、情報ごとの重要度は考慮せず、設定内容に合致する情報をそのまま提示するだけであるため、ページ閲覧者の要求を反映できない場合がある。

6 おわりに

本研究では、ページ閲覧者の嗜好に合わせ、Web ページを動的に生成し、提示するシステムの実現を目的とし、個人化を考慮したデータ格納モデルと、視聴履歴に基づく、個人化した Web ページの生成手法を提案した。また、そのプロトタイプシステムを実現した。

提案モデルに基づき、コンテナやコンテンツ、リレーションといった素材単位のデータを用意し、提案手法を用いて、視聴履歴からページ閲覧者の視聴傾向を抽出して利用することで、各ページ閲覧者の視聴目的や興味に適した Web ページを、ページ閲覧者に必要な情報だけで生成できる。

今後は、まず、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、評価値 v の導出に適した式を考察することや、視聴傾向として、リレーションの全く同じ辿り方のみを考えるのではなく、ノイズを含む大雑把な視聴傾向を検出して、個人化した Web ページを生成する方法も考えていく。さらに、Web デザインの知識を踏まえ、コンテンツ間の関係を考慮した、ページレイアウトの決定方法をより深く考え、同時に、様々なクライアント環境に対するページレイアウトの決定方法も考えていく予定である。そして、最終的には、懐徳堂コンテンツの個人化提示システムを実現し、実証実験を行う予定である。

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 (JSPS-RFTF97P00501)」と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懐徳堂研究」(課題番号 13309011) によっている。ここに記して謝意を表す。

参考文献

- [1] CAST: “CAST: Center for Applied Special Technology,” <http://www.cast.org/>
- [2] World Wide Web Consortium: “Extensible Markup Language (XML) 1.0 (Second Edition),” <http://www.w3c.org/TR/REC-xml/>
- [3] Florian, M., and Andrea, L.: “Cheese: Tracking Mouse Movement Activity on Websites, a Tool for User Modeling,” Proc. CHI 2001., Seattle (April. 2001).
- [4] HYPERGENIC: “Hypergenic WebNize 1000 Metro,” <http://www.hba.dk/metro/>
- [5] Motoki Miura, Buntarou Shizuki, and Jiro Tanaka: “inlineLink: Inline Expansion Link Methods in Hypertext Browsing,” Proc. International Conference on Internet Computing (IC2001), Vol. II, pp. 653–659 (June. 2001).
- [6] Nikkei Business Publications: “IT Pro (My Page),” <http://itpro.nikkeibp.co.jp/members/mypage/index.jsp>
- [7] World Wide Web Consortium: “Resource Description Framework (RDF) Model and Syntax Specification,” <http://www.w3c.org/TR/REC-rdf-syntax/>
- [8] WORLD WIDE WEB consortium (W3C): “XML Linking Language (XLink) Version 1.0,” <http://www.w3c.org/TR/xlink/>
- [9] WORLD WIDE WEB consortium (W3C): “XSL Transformations (XSLT) Version 1.0,” <http://www.w3c.org/TR/xslt/>
- [10] Yahoo Japan Corporation: “My Yahoo!,” <http://my.yahoo.co.jp/>
- [11] 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).
- [12] 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).