

XML データベースからの動的 Web ページ生成環境における 変更検出・通知方式

宮崎 慎也[†] 馬 強^{††}
角谷 和俊^{†††} 田中 克己^{†††}

HTML のみによる静的な Web ページに対し、アプリケーションサーバ、データベースなどを用いて動的に生成される Web ページが増加している。例えばサイト内部のデータを対象にサイト内検索機能を提供する Web サイトなどが多くみられ、こうした環境下では内部データを対象に変更通知を行う必要がある。本稿では、こうした動的な Web ページ生成環境における変更検出、通知方式に関する提案を行う。Web サーバ内において、動的な Web ページ生成のための素材となる各コンテンツの変更の構造・内容解析を行い変更の評価を行う。また、変換プログラムに着目することで、生成されるページスタイルやレイアウトなどの変更を通知することが可能である。このように動的な Web ページ生成環境下で、データとスタイルの両者の変更解析を行うことで、サーバ内部のデータ変更を自動的にユーザへ変更通知することが可能である。

Change Detection and Notification of Dynamic Web Page Generation Environments Using XML Databases

SHINYA MIYAZAKI,[†] MA QIANG,^{††} KAZUTOSHI SUMIYA^{†††}
and KATSUMI TANAKA^{†††}

Many Web sites, especially the B2B and B2C sites, generate their Web pages from database dynamically. In this paper, we propose a method for analyzing and notifying changes of these dynamic Web pages. In our method, the changes are detected by monitoring the modifications of source database and Web page's generation functions. For picking out the important changes to notify users, the changes are analyzed based on 1) content and structure of source data, and 2) style and layout specification for generating Web pages.

1. はじめに

近年、Web 上の情報は大量かつ複雑なものとなっている。巨大な Web サイトでは、頻繁にページの変更、追加が行われ、これら大量の変更の中から、ユーザが興味ある情報を常に獲得することは困難である。また、情報を提供する Web サイト側では、自身の新しいコンテンツをユーザに通知したいという要求が考えられる。そのため、これまでもさまざまな Web の変更監視・通知方式に関する研究¹⁾²⁾³⁾ がなされてき

たが、以下のような問題点が考えられる。

- 従来の研究の多くは、HTML ページあるいはページ内の一部を対象にしており、動的なページ生成環境における内部データを考慮していない。
- ページのスタイルの変更を考慮していない。

現在、CGI、ASP、データベースなどの利用が進み、動的なページ生成を行う Web サイトが増加しており、本論文ではこうした動的な Web ページ生成環境における、ページ生成の元となるサーバ内部のデータと、生成されるページのスタイルの両者を考慮した変更通知手法を提案する。通常、ユーザが Web の変更を知るために自動巡回ツールなどを用いることがあるが、動的にページが生成される環境では当然ながら内部のデータソースにアクセスすることはできない。本研究では、このような動的ページ生成の素材となるサーバ内部のデータと、生成されるページスタイルに着目する。

[†] 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 System Engineering, Graduate School of Science and Technology, Kobe University

^{††} 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 Science, Graduate School of Science and Technology, Kobe University

^{†††} 京都大学大学院情報学研究科社会情報学専攻

Division of Social Informatics, Graduate School of Informatics, Kyoto University

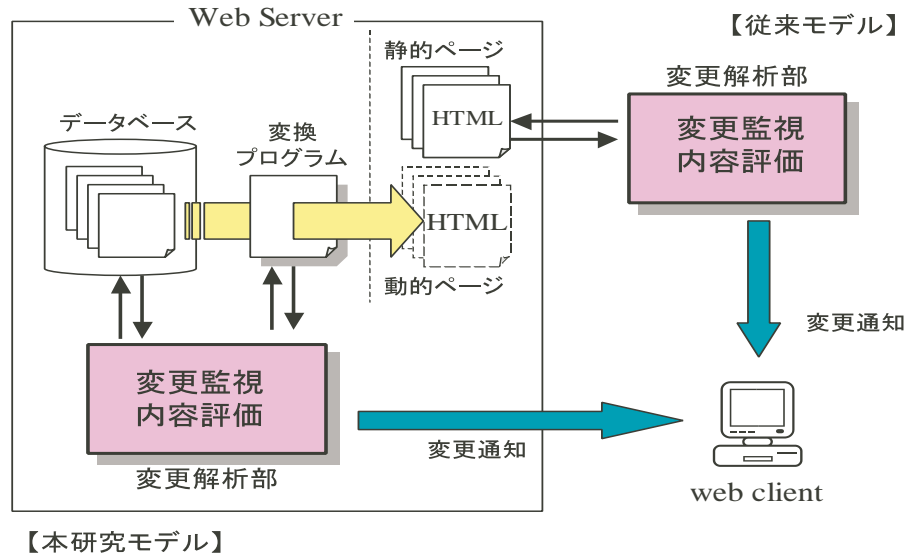


図 1 変更通知システムのモデル

我々の以前の研究¹⁾も、Web サイトに公開されている HTML ページを対象とした変更通知機構であった。そのモデルと本論文で考えるモデルをあわせて図 1 に示す。Web サイト内には静的なページと動的に生成されるページの両者が存在する場合が考えられ、従来のモデルでは外部から Web サイトにアクセスし取得できる HTML ページのみを対象としていた。つまり静的な Web ページを対象にした変更通知モデルを提案したが、本論文で考えるモデルは、動的な Web ページ生成環境での変更通知に焦点をあて、サーバ内部のデータを対象とした変更解析、通知手法である。

外部から Web サイトにアクセスし、公開されている Web ページ (HTML 文書) に対して監視、解析する従来の変更通知モデルの場合、外部からアクセスできないページや、E コマースサイトのようにユーザのアクセス履歴やアクセス状況に応じてデータベースから動的に生成されるページでは、変更の検出、評価が困難である。そのためサーバ内部のデータと、ページ生成のための変換プログラムに焦点をあてた変更監視、評価、通知システムを提案する。

ページを構成するデータとページ生成のための変換プログラムを解析することで、最終的に生成されるページの更新を捉える。また、ユーザのアクセス履歴、行動履歴などを組み合わせることで、サーバ側としては自身の新規情報をユーザへ自動通知、個別化された変更通知を行うことができ、またユーザ側では、個別化された重要情報の自動獲得が可能であると考えられる。

ここで本論文の特徴を以下にまとめる。

- 動的に生成されるページを対象とした変更検知と評価方法の提案。

想定する動的ページ生成環境では、ページの元となるソースデータと変換プログラムにより動的にページが生成される。ソースデータと変換プログラムの両者を対象に変更検知と評価を行う。評価すべき変更は、データベース内の実際の変更データの集合と変換プログラムの解析により得られる監視すべき変更の集合の積として与えられる。

- XML データに対する構造と内容を考慮した変更評価。

本論文ではページ生成の元となるソースデータとして、XML⁶⁾ データを想定する。XML データの構造と内容に基づいてその変更*を評価し、評価に基づいた変更通知を行う。

- 変換プログラムによるページのスタイル、レイアウトを考慮する。

ページ生成のための変換プログラムから、生成されるページのスタイル、レイアウトの変更を検出する。変換プログラムの一部として XSLT⁷⁾ ドキュメントを想定し、これを解析することでページ生成の素材として用いられるデータを特定する。さらに XSLT ドキュメント自身の変更を解析することで、ページスタイル、レイアウトの変更を検出、通知する。

*本論文では、文書単位でのデータの追加を変更として扱う。

2. 関連研究

WebCQ²⁾ は Web ページ内の変化を発見し、ユーザにパーソナライズされた通知を行うシステムである。WebCQ の特徴は、監視・追跡する変更の種類の豊富さと、ユーザごとのパーソナライズされた変更情報の呈示である。WebCQ の監視対象はユーザが指定する任意の Web ページであり、内部データや変換プログラムなどの動的ページ生成環境を特別考慮したものではない。また変更通知はユーザの興味に基づいて行われるため、ユーザの興味は明確なものである必要がある。我々の研究のアプローチでは、ユーザの明確な興味を必要とせず、変更に対する意味的な評価から重要と推定される情報の通知を行う。

Lim ら³⁾ は SCD (Semantic Change-Detection) という二つの HTML 文書間の意味的差異を検出するアルゴリズムを提案している。SCD では二つの文書間のデータの差異を意味的な構造として抽出、呈示する。本研究では一つの文書を、他の複数の文書と比較することで、その重要性を推定するものである。また、本研究では動的ページ生成環境を想定しているため、ページの素材となるデータと変換プログラムの両者によってその変更を捉える。

小島ら⁴⁾ は Web 上の複数のコンテンツに対して、半構造データの値、構造の差異を発見する研究を行っている。複数の半構造データからボトムアップに DataGuide オブジェクトを作成し、データの用語、構造の差異から、データ間の共通スキーマをツリー形式で生成する。本研究で対象とする XML ドキュメントも半構造データであり、変更評価の際の比較対象を特定する手法として有効であると考えられるが、本論文の主目的は重要変更を推定し、ユーザに通知することであり、本論文においては文書の構造とノード名からより簡易的に比較対象を特定する手法を用いた。

3. XML データの変更解析

Web サーバ内部に格納されている XML 文書に対する変更解析について述べる。本研究のコンセプトはあらかじめユーザによって指定された変更だけを通知するのではなく、変更を解析、評価した結果、重要と推定される変更をユーザに通知するものである。そこで、追加された XML 文書に対して、すでに存在する過去の XML 文書との構造、内容の比較を行う。

XML 文書の構造を明示的に指定する場合、おもに DTD や XML Schema などが用いられる。共通の構造定義に従う XML 文書では、文書間の構造は共通であ

るが、その場合でも、文書中に用いられるタグの中には、存在してもしなくてもよい要素や、出現回数が自由な要素が存在するなど、文書単位での構造の差が存在する可能性がある。また、本論文で考えるデータ群は共通の構造定義をもつ XML 文書ではなく、異なる構造定義をもつ文書群を対象に、追加されたドキュメントの構造、内容の特徴を評価する。

以下、追加文書の各テキストノードに対し、過去の文書群と比較、評価するために、3.1 では、過去の文書中の同一の種類の情報を記述したノードを特定する方法について述べる。また 3.2 では追加文書と過去の文書群との構造的な差異について、3.3 では内容の差異について述べる。3.4 では、追加文書の構造の差異、内容の差異の両者から、変更評価を決定する手法について説明する。

3.1 比較ノードの特定

XML 文書は、その構造に意味があり、また一般に意味付けされたタグを用いて情報記述される。追加された文書内の内容を、過去の文書と比較する場合、当然同一のタグで囲まれた情報が過去の文書中に存在する場合は、それらの情報と内容比較を行うことが好ましい。

そこで、まず新しく追加された XML 文書内のすべての内容 (テキストノード) を含む要素タグに関して、過去の XML 文書に対する問い合わせを行う。図 2 を例にとると、追加された文書の構造から、内容 (テキストノード) を伴う A, B, C の各要素に対して、それぞれのノードの XPath⁸⁾ 指定

- `/root/a/c/A`
- `/root/b/B`
- `/root/b/C`

を抽出し、それぞれのパスを問い合わせに用い、過去の XML 文書に対してパス検索を行う。たとえばノード A によって囲まれた内容 (テキストノード) は、過去の XML 文書中のパス `/root/a/c/A` によって指定される内容と比較、評価を行うことが望ましいと考えられる。

しかしながら、本論文では共通のスキーマを持つ XML 文書だけを対象としている訳ではなく、文書構造が異なる二つの文書間でも、共通の情報を記述している場合が考えられる。つまり、完全にパス指定が一致するノード以外も比較対象となるノードが存在する可能性があると考えられる。そこで今回は、同一情報が記述される比較ノードの特定を次のように決める。

- (1) 対象ノードのパス指定と、同様の指定で抽出されるノード

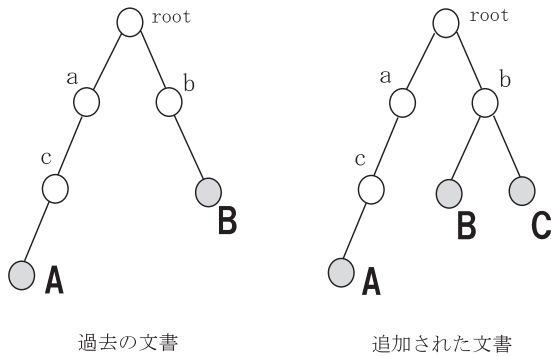


図2 XML文書の構造例

- (2) 対象ノードのパス指定において、末端のノードより上位に存在するノードを含み、かつノード名が共通するようなノード
(各ノードの前後関係は保持する.)
- (3) 対象ノードのパス指定において、末端のノードより上位に存在する各ノードの一部から形成され、かつノード名が共通するようなノード
(各ノードの前後関係は保持する.)

図2右部のノードAを例にとると、

- (1)では、“/root/a/c/A”よって指定されるノード
- (2)では、“/root//a//c//A”、“*/root//a//c//A”などの問い合わせで検索されるノード
- (3)では、“/root/a/A”、“/root/c/A”、“/root/A”、“/a/c/A”、“/a/A”、“/c/A”などで指定される各ノード

が、それぞれ比較ノードとして選択される。

3.2 構造の差異

ほとんどの文書には表れない種類の情報が記述された文書は、特徴を持っていると考えることができる。3.1から抽出される各パス検索の結果、それぞれの要素に関して、過去のどれぐらいの文書で同様のノードが見られるかということに注目する。

ここで図2を例にとって説明すると、A、Bの要素に関しては、過去のXML文書内に存在し、また要素Cに関しては、過去のXML文書には登場していない。このように、過去の文書にはほとんど見られない種類の情報を記述したノードが発見された場合、新しく追加された文書には構造上の特徴があると考えられる。また検索の結果、同種の情報を記述したと推定されるノードが過去のXML文書から得られた場合でも、それらが内容を伴わない空要素であった場合も、新しい種類の情報の追加と考えることができる。

ここで、総XML文書数を n 、新規文書内のある要

素ノードAの過去の文書に対するパス検索結果数(該当文書数)を m_A とすると、ノードAの文書構造上の特徴量 $S(A)$ を、

$$S(A) = 1 - \frac{m_A}{n} \quad (1)$$

とする。また、空要素タグが検索された場合は、検索結果には含めない。

3.3 内容の差異

追加されたXML文書の内容を過去の文書と比較し、評価する。様々な形式のデータが考えられるが、本研究では、おもに数値データとその他のテキストデータに分けて内容を評価する。3.1に示した方法で特定される比較ノードを対象に、それぞれの内容評価を行う。

3.3.1 数値データ

株価や商品の価格など、何らかの数値の情報を記述している場合、過去の文書中の情報に対して大きく違っている場合ほど、変更通知を行う必要性が高いと考えられる。そこで、追加された文書中の数値の、過去の文書中の各数値に対する偏差値 $h(i)$ を評価の尺度に用いる。3.1で述べた比較ノードが数値データでない場合は、そのノードは比較対象から除く。

追加文書中のテキストノード i (数値データ)の内容評価値は、

$$V(i) = \left| \frac{h(i)}{50} - 1 \right| \quad (2)$$

とする。これは、対象とするデータが、過去の文書中の比較対象なるデータの平均的な値に対し、大きく離れているほどその評価値は大きくなる。

3.3.2 テキストデータ

ニュースコンテンツのようにテキストを主な情報として記述している場合を考える。ここで内容の評価に関して、我々は以前の研究¹⁾において用いた、馬ら⁵⁾によって提案された新鮮度、流行度等の尺度を用いる。文書中のテキストノード i に関する新鮮度 $V_f(i)$ 、流行度 $V_p(i)$ はそれぞれ、

$$0 \leq V_f, V_p \leq 1 \quad (3)$$

の範囲で算出される。その定義を以下に示す。

$$\begin{aligned} V_f(i) = & \alpha * freshnum(i) + \beta * freshcd(i) \\ & + \gamma * freshae(i) + \sigma * freshid(i) \quad (4) \\ \alpha + \beta + \gamma + \sigma = & 1.0, \alpha, \beta, \gamma, \sigma \geq 0 \end{aligned}$$

ここで、 $\alpha, \beta, \gamma, \sigma$ はそれぞれ重みであり、また ω は過去の文書中の各比較ノードにおける、ノード i に類似するノードの集合である。またそれぞれ、

$$freshnum(i) = \frac{1}{\log_2(2+m)}$$

$$fresh_{cd}(i) = \log\left(\frac{1}{m} \sum_{j=1, b_j \in \omega}^m \left(1 - \frac{v(i) \cdot v(b_j)}{|v(i)| |v(b_j)|}\right)\right)$$

$$fresh_{de}(i) = \log_2 \frac{n}{m}$$

$$fresh_{td}(i) = \log\left(\frac{1}{m} \sum_{j=1, b_j \in \omega}^m (t(i) - t(b_j))\right)$$

と与えられる。\$n\$ は比較ノードの総数、\$m\$ は各比較ノードのうち、ノード \$i\$ と類似するノード \$b_j\$ の総数、\$\omega\$ は類似ノード \$b_j\$ の集合、\$v(i)\$ はノード \$i\$ から生成した特徴ベクトル、\$t(i)\$ はノード \$(i)\$ を含む文書の追加時間である。

$$V_p(i) = e^{\lambda_1 d} + e^{-\lambda_2 t_d} \quad (5)$$

ここで、\$\lambda_1 (> 0)\$、\$\lambda_2 (> 0)\$ は重み、\$d\$ は類似ノードの密度、\$t_d\$ は \$i\$ と各類似ノードとの平均時間距離である。

追加文書中の内容を含む各要素ごとに新鮮度、流行度を計算する。図 2 の場合、各要素 A、B、C の新鮮度 (\$V_f(A)\$、\$V_f(B)\$、\$V_f(C)\$)、流行度 (\$V_p(A)\$、\$V_p(B)\$、\$V_p(C)\$) をそれぞれ求める。ここで、要素 C のように、過去の文書中に同一の要素が存在しない場合、つまり比較対象が存在しない場合は、

$$\begin{cases} V_f = 1 \\ V_p = 0 \end{cases} \quad (6)$$

とする。

また XML 文書では、各要素に属性情報を付加することが可能であるが、本稿ではこれら属性情報については、考慮していない。

3.4 変更評価

3.2 にて述べた構造上の差異、3.3 にて述べた内容の差異の両者を考慮して、追加文書の評価を決定する。ここで文書内の各ノードの内容評価値に対し、ある一定以上の構造上の特徴量有する場合、その特徴量を重みとして付加する。ここで構造上の特徴量を重みとして付加するかどうかの閾値 \$\alpha\$ (\$0 < \alpha \le 1\$) を用いる。

追加文書中のあるテキストノード \$i\$ の変更評価値 \$V(i)\$ (数値データ)、\$V_f(i)\$、\$V_p(i)\$ (テキストデータ) は内容評価値に構造上の特徴量を重みとして付加し、以下に再定義する。

$$V(i) = \begin{cases} S(i) * V(i) & (\alpha \le S(i) \le 1) \\ \alpha * V(i) & (0 < S(i) < \alpha) \end{cases} \quad (7)$$

$$V_f(i) = \begin{cases} S(i) * V_f(i) & (\alpha \le S(i) \le 1) \\ \alpha * V_f(i) & (0 < S(i) < \alpha) \end{cases} \quad (8)$$

$$V_p(i) = \begin{cases} S(i) * V_p(i) & (\alpha \le S(i) \le 1) \\ \alpha * V_p(i) & (0 < S(i) < \alpha) \end{cases} \quad (9)$$

これら構造上の特徴抽出や、数値、テキスト等の内容評価などを利用し、さまざまな変更通知が可能であると考えられる。たとえば、大きな構造上の特徴を有するデータが追加された場合のみ通知する、あるいは数値の変動やテキスト内容の特徴量に応じて通知のタイミングを決定するなど、各ユーザに個別の変更通知が可能であると考えられる。

4. 変換プログラムの解析

XML で記述されたコンテンツデータから、動的に Web ページを出力するための変換プログラムを解析する。一般に、動的なページ生成には ASP、CGI などを利用した変換スクリプトが考えられ、また最終的なページの生成に XSLT の利用が考えられる。本論文では、ASP ファイルなどに記述されたデータ抽出のための変換スクリプト、抽出されたデータから最終的なページに変換、スタイル付けを行う XSLT など、データベースに蓄えられたデータから HTML ページを生成するまでの処理を統合したものを変換プログラムとして扱う。本節では、変換プログラムに対する解析の一例として、特に XSLT ドキュメントを対象とした解析について述べる。

動的に生成されるページの変更は、ページの元となるソースデータの変更と、ページ生成のためのデータを指定する変換プログラムの両者によって決定される。例えば、XML で記述されたデータが新規追加された場合でも、そのデータが変換プログラムによって最終的に Web ページとして出力されなければ、ユーザ (client) に通知する必要性はない。そこで、データベースへの問い合わせ式や XSLT ドキュメントを解析することで、最終的に XSLT によって生成されるページ中に用いられる XML 文書、また文書中の部分を特定することができる。図 3 に簡単な XSLT ドキュメントの記述例を示す。

XSLT ドキュメントを解析することで、どの XML 文書が参照されるか、また文書内のどの部分が利用されるかを特定することができる。これには XSLT 内部で利用される、

- <xsl:template>

```

<?xml version="1.0" encoding="Shift_JIS"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:template match="/">
  <html>
  <head><title>・・・</title></head>
  <body>
    <div>
      <xsl:value-of select="document('file1.xml')//c/A"/>
      <xsl:value-of select="document('file2.xml')//b/C"/>
      ...
    </div>
    <xsl:apply-template select="root/a/c/A"/>
  </body>
</html>
</xsl:template>

<xsl:template match="root/a/c/A">
  <h1>
  <xsl:value-of />
  </h1>
</xsl:template>
...
</xsl:stylesheet>

```

図 3 XSLT ドキュメントの例

- <xsl:value-of>
- <xsl:for-each>

などのノード指定をともなう各要素に着目する。<xsl:template> 要素ではおもにカレントノードが指定され、<xsl:value-of> や <xsl:for-each> などの要素では、select 属性における XPath 指定によってコンテキストノードが指定される。カレントノードとコンテキストノードの組み合わせから、指定されるデータを特定する。

また、適用される XSLT ドキュメント自身が変わった場合、生成されるページのスタイル、レイアウトの変更と捉えることができる。ここで、変更されたスタイルをユーザへ配信されるの通知情報のスタイルとして利用し、スタイルの変更を視覚的に表現することで、効果的な変更通知を行うことができると考える。

5. 変更通知システム

現在、プロトタイプシステムを開発中である。XML 文書を格納するデータベースに Yggdrasil 1.5⁹⁾、Web サーバに IIS 5.0 を使い、Windows2000 上で開発中である。変更通知システムを含む Web サーバ内の構成図を図 4 に示す。XML データベースへの文書追加と XSLT ドキュメントの変更を監視する変更監視部、変更に対し解析、評価を行う変更解析部、解析結果に応じて通知情報を作成、配信する変更通知部から構成されている。

このシステムが対象とする動的ページ生成環境の変換プログラムは ASP ファイルと XSLT ドキュメント

によって構成される。ユーザの ASP ページへのアクセスにより、XML データベースへの問い合わせが実行され、問い合わせ結果が一つの XML 文書として返され、指定される XSLT ドキュメントにより HTML に変換、表示される。XML データベースへの問い合わせ、対応付けられる XSLT ドキュメントの指定は ASP ファイル内に記述される。ページ生成までの流れを以下に示す。

- (1) ASP ページに記述されている XML データベースへの問い合わせ (XPath 指定) の実行。データベース内の複数文書に対し、指定された XPath による問い合わせを行う。
- (2) XPath 指定に該当するノード部分が、データベース内の各ドキュメントから抽出され、仮想のルートノードに結合された、一つの仮想 XML ドキュメントが生成される。
- (3) 生成された XML ドキュメントが、ASP ファイル内に指定された XSLT ドキュメントによって HTML に変換、ユーザへ呈示される。

変更解析部では、XML データベース内に格納される各 XML ドキュメント、ページ生成に利用される ASP ページ、XSLT ドキュメントを対象に変更監視を行う。

5.1 動作概要

データベース内部の変更 (XML ドキュメントの追加) を検出した場合の、変更通知までの流れを説明する。

- (1) 各 ASP ページに記述されている XML データベースへの問い合わせを解析し、追加された XML ドキュメントがページ生成のために利用されるかデータを含んでいるかどうか判定する。
- (2) データベースへの問い合わせにより追加ドキュメントから抽出されるノードを特定し、ASP ファイル内で指定されている XSLT ドキュメントの解析を行う。XSLT ドキュメントの解析により、最終的に生成されるページに利用されるノードを特定する。
- (3) 特定されたノードに対し、構造上の特徴量、内容評価を行い、変更評価値を算出する。
- (4) あらかじめ設定された閾値を元に、変更評価値が閾値を上回った場合、通知情報として選択される。

複数の変更に対し、それぞれ上記のステップが実行され、通知する情報 (ノード) が最終的に決定される。

5.2 変更通知

変更通知は、通知情報を記述したページを自動作成し、プッシュ型配信機構によって各ユーザに配信、呈

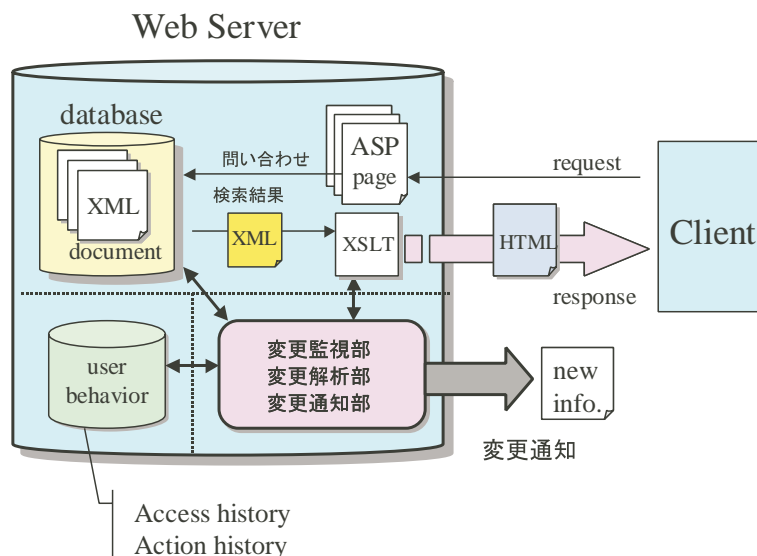


図4 構成図

示することで行われる。

通知する情報として選択されたノードは、XSLT ドキュメントによって HTML による通知情報に変換され、生成されるページと同じスタイルで通知される。作成される通知情報が複数ページに及ぶ場合、これら各通知情報に対するリンク集ページ作成し、配信する。

変更通知を個別化する方法として、サーバ側で個別化する場合とクライアント側での個別化する場合の二通りが考えられる。サーバ側では、各ユーザのアクセス履歴、行動履歴に応じて通知すべき変更を選定することで、あらかじめ個別化された通知コンテンツを各ユーザに配信する。また、クライアント側で通知の個別化を行う方法としては、各ユーザに同一の通知コンテンツを配信し、クライアント側の各ユーザのユーザプロフィールによって通知コンテンツを変換、呈示する手法が考えられる。これは、通知コンテンツを XML で記述し、ユーザプロフィールを XSLT ドキュメントで記述することで実現する。

6. おわりに

本論文では、動的な Web ページ生成環境における変更検出・通知方式について述べた。動的生成ページにおける変更は、ページの素材となるデータの変更とページ生成のためのデータを指定する変換プログラムの両者から検出される。従来の Web の変更監視、通知に関する研究に対して、新たな特徴として

- ページ生成の素材となるコンテンツデータと、ペー

ジ生成のための変換プログラムのそれぞれを対象データとする。

- 変更データに対する、内容評価、構造上の特徴抽出。
 - Web ページのスタイルの変更を取り扱う。
- などの点を挙げる。

また、今後の検討すべき課題として

- XML 文書中の属性情報の扱い。
- ユーザのアクセス履歴、行動履歴からユーザの興味、嗜好を抽出し、通知コンテンツを個別化する手法の確立
- 異なる構造を持つ文書間における、ノードの内容比較方法の改善。

などを挙げる。また、変換プログラムは動的にページを生成するためのルールであり、この生成ルール自体を対象とした変更検出・通知方式などを考えている。

謝 辞

本研究の一部は、文部科学省科学研究費「分散型ハイパーメディアからの構造発見とアクセス管理」(課題番号 12680416)、「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 13224054) の援助を受けており、また、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) によっております。ここに記して謝意を表すものとします。

参 考 文 献

- 1) 宮崎慎也, 馬強, 田中克己: WecSCAN: Web サイトの変更発見と放送型変更通知, 情報処理学会論文誌: データベース Vol.42 No.SIG10(TOD10), pp96-107(2001).
 - 2) Ling Liu, Calton Pu, and Wei Tang. WebCQ-detecting and delivering information change on the Web. In *Proc. of CIKM'00* (2000).
 - 3) Seung Jin Lim, Yiu-Kai Ng: An Automated Change Detection Algorithm for HTML Documents Based on Semantic Hierarchies. ICDE 2001: 303-312
 - 4) 小島岳史, 清光英成, 田中克己: 半構造データの差異発見・比較のための共通スキーマ生成機構, 情報処理学会研究報告, Vol. 2001, No. 70, pp. 249-256 (2001)
 - 5) 馬 強, 角谷和俊, 田中克己: 放送型情報配信システムのための時系列性を考慮した情報フィルタリング, 情報処理学会論文誌データベース (TOD7) (2000).
 - 6) W3 consortium. Extensible Markup Language(XML)1.0. <http://www.w3.org/TR/REC-xml/>
 - 7) W3 consortium. XSL Transformations(XSLT)1.0. <http://www.w3.org/TR/xslt/>
 - 8) W3 consortium. XML Path Language(XPath)1.0. <http://www.w3.org/TR/xpath/>
 - 9) 株式会社メディアフュージョン. <http://www.mediafusion.co.jp/>
-