

XML ページの分解・再構築技術¹

3P-12

浜田伸一郎 関俊文

株式会社 東芝 研究開発センター

1 はじめに

急速に成長している WWW において、新技術である XML¹が熱い注目を浴びている。本稿では、異なる複数のウェブサイトから配信される XML ページを分解して収集・整理し、さらに一つのページとして統合するクライアント側の情報統合技術を提案する。この技術は XML に含まれる情報の再利用性・可搬性を高め、かつクライアント中心のネットワークアーキテクチャを実現する。

2 背景

XML は、SGML を WWW 上で交換する目的で簡易化された汎用マークアップ言語であり、HTML と同様ハイパーメディアとしての利用が想定されている。しかし HTML とは異なり、内容に関する定義と整形方法に関する定義は、別モジュールとして陽に分離されており、それら分離されたモジュールが関連付けられて、一つのページとして提供される。その特徴の恩恵の一つとしてページ内容の再利用性がある。同じページ内容でも異なる整形方法を提供したい場合、HTML では、内容と整形方法の記述が混在しているため、内容記述も含めてページ全体を再作成しなければならない。一方 XML では基本的に、それに対応するスタイルシートを適用し直すことでページ内容を再利用できる。

しかし、XML における再利用単位はページ単位であり、より小さな単位での再利用が望まれる場合がある。たとえば商品カタログなどを列挙する XML ページを提供するウェブサイトは大きなデータベースを抱えており、クライアントに対して、すべてのカタログを一つの XML ページとして提供することは不可能である。したがって何らかのフィルタを用いて、適当なカタログの集合として XML ページを提供することになる。しかしその場合、商品カタログの組み合わせを変更するたびに、クライアント側は XML の内容モジュール全体を、ダウ

ンロードし直さなければならず、高いアクセスコストを強いられる。

また複数のウェブサイトから、横断的に情報を閲覧したい場合がある。例えば、異なるパソコンショップサイトで扱っているパソコンを比較したいという場合に、現状ではそれぞれのサイトのウェブページを同時に開き、ユーザが目で見比べて比較しなければならない。このようにウェブサイトをもたぐると、電子的なメリットを全く享受できなくなる。

3 XML ページの分解と再構築

本稿では前節の問題を解決するため、XML ページから、再利用可能な部分文書(以下このような文書を「再利用文書」と呼ぶ)を切り出し、これら取り出された再利用文書の任意の組み合わせによって、XML ページをクライアント側で再構築するシステムを提案する。再構築される XML ページは、構成する再利用文書が属していた元の XML ページのスタイルシートを適用することにより表示する。しかしその際、スタイルシートは特定の文書構造を期待しているため、スタイルシートが適用されていた元の XML 文書と同じ文書構造を忠実に再現する必要がある。

また XML ページをクライアント側で再構築するには、元の XML ページで用いられている画像ファイルやスクリプトファイルなどの外部モジュールが必要となる。したがって、それぞれ再利用文書が参照している外部モジュールをすべてクライアントのファイルシステムにミラーリングする必要がある。

3.1 記述ルール

再利用文書を取り出すことのできる XML 文書には、①再利用文書の領域を示すフラグ、②コンテンツ文書において再利用文書が挿入される要素の位置を示すフラグ(以下これを「接続位置フラグ」と呼ぶ)、および③外部モジュールへの参照を示すフラグ(以下これを「外部モジュール

¹ Decomposition and Reconstruction Technology for XML pages
Shinichiro HAMADA(shinichiro.hamada@toshiba.co.jp) Toshiyumi SEKI(toshiyumi.seki@toshiba.co.jp)
Research and Development Center, TOSHIBA Corporation
Yanagi-cho 70, Saiwai-ku, Kawasaki City, Japan

フラグ」と呼ぶ)に関する記述ルールを施した。図1の例のように、これらフラグは、それぞれ属性 component、container、external によって示される。

また XML ページを再構築するために必要な、再利用文書・コンテナ文書・スタイルシート・および外部モジュール名の情報を維持する XML モジュールを「XML オブジェクト」と呼び、図2のような形式で記述する。

3.2 XML ページの再構築

XML ページの再構築は次の手順で行う。まず対象とする XML オブジェクトそれぞれに含まれるスタイルシートのいずれを適用するかを決定する。次に選択したスタイルシートを含む XML オブジェクトが持つコンテナ文書と再利用文書を組み合わせた XML 文書を、これから再構築する XML ページのベースとする。続いて、接続位置フラグで示される文書要素に対してその他の XML オブジェクトに含まれる再利用文書をすべて挿入する。最後に、外部モジュールフラグを手がかりとして XML 文書が参照している外部モジュールをすべて探し出し、適切なディレクトリにコピーする。これらの作業により、対象の再利用文書がすべて含まれる XML ページが実現できる。

4 本技術の応用

本技術を用いたプロトタイプシステムを開発した。本システムは IE5 上で動作し、XML ページ内部のスクリプトプログラムと、エクスプローラバーとして実装されているアプリケーション部から構成される。図3はその実行画面例であり、右側にウェブページ・左側にエクスプローラバーが表示されている。

ユーザは、XML ページ上に列挙されている任意のカタログをドラッグ&ドロップ操作によって、アプリケーション部に XML オブジェクトとして取りこむことができる。今までに取りこまれた XML オブジェクトはアイコンとしてペイン上にリスト表示され、そのうちフォーカスされているアイテムについては、プレビューウィンドウにおいて整形表示される。これらのアイテムは、フォルダーを用いて整理したり、削除・検索などを行うことができる。

また XML オブジェクトをすべて含む XML ページを生成して、ウェブブラウザ上に表示することができる。このとき複数のスタイルシートがある場合は、候補の中から選択することができる。また自分のローカルファイルシステムにオリジナルのスタイルシートがある場合、それを指定して適用することも可能である。

5 おわりに

本稿では、XML ページを XML オブジェクトに分解し、自由に組み合わせて XML ページを再構築する技術を提案した。今後の課題としてより柔軟な XML 部分文書の組み合わせの実現を考えている。

参考文献

- ¹ W3C: Extensible Markup Language(XML) 1.0,
http://www.w3.org/TR/REC-xml

```
<doc> <title>catalogs</title>
<cats container="yes">
  <motor component="yes">
    <outline external="yes">motor1.jpg</outline>
  </motor></cats></doc>
```

図1 カタログ文書のサンプル

```
<catalog>
  <cats container="yes"><motor component="yes"> ...
</site-info>
  <external style="xsl">cat.xsl</external>
  <external>motor1.gif</external>
</site-info></catalog>
```

図2 XML オブジェクトのサンプル

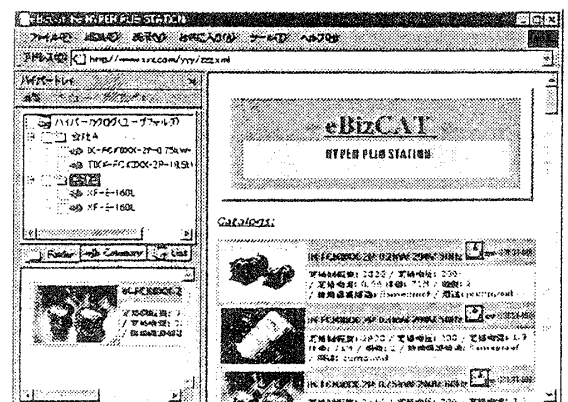


図3 アプリケーション実行画面例