

O_009

テンプレートを利用した分散 XML データの共有システム A System for Sharing Distributed XML Data Using a Template

服部 哲†
Akira Hattori 田畠 邦晃†
Kuniaki Tabata

1. はじめに

Web 技術を利用して組織間で情報を共有することは重要であり、今日では、XML(eXtensible Markup Language)を活用することが重要になっている[1]。XML はタグを利用してデータを記述する言語であるが、タグセット(タグの名前と構造)を自由に決め、データの構造と意味を保持してインターネット上でデータ交換できる点に特徴がある[2]。各組織はそれぞれの視点で集めた情報を XML で分かりやすく記述することができるため、各組織に分散した XML データを共有するためのシステムの開発が期待されている。

分散した XML データを共有するためのシステムに関する従来研究は大きく二つに分類できる。一つは、XML データの統合のためにメディエータを導入するものである[3-6]。メディエータは各組織が利用するタグセットと特定分野における共通のタグセットやオントロジーとの対応関係を管理し、ユーザからの要求に応じて分散した XML データを検索し、結果をタグの対応関係に基づき統一のタグセットに変換してユーザに提供する。すべての XML データのタグを考慮して共通のタグセットやオントロジーを作成することは困難であるため、この方式では、対応関係を定義できないタグで記述されたデータを共有できないという問題がある。

もう一つは、各組織が管理するサーバの名前やインターネット上のアドレスなどの情報を管理するディレクトリを導入するものである[7-10]。この方式は、各組織がその組織で管理するデータを共通のタグセットを利用して記述することを前提としている。しかしこの方式の場合も、共通のタグで記述できないデータを共有できないという問題がある。また、どちらの方式においても、各組織の XML データを組み合わせ、新たな XML データを作成するというような、データの再利用は困難である。

ところで、同じ分野の情報を収集している組織では、ある程度項目は共通していると考えられる。そこで本研究では、ツリー構造の上位のタグセットをテンプレートとして組織間で共通にし、各組織はそれより下位のタグを自由に追加できるというアプローチを探る。このアプローチでは多様な視点に基づく組織独自の情報も記述し容易に交換できる。本論文ではこのアプローチに基づき、分散した XML データを共有するためのシステムを提案する。

2. テンプレートベースのアーキテクチャの提案

2.1 アプローチと意義

XML の特徴の一つはタグの拡張性、つまり XML の利用者が自由にタグセットを決められる点にある。そこで、情

報共有に係わる組織がどの程度自由にタグを利用できるかに注目すると、次の三つのレベルが考えられる。

レベル 1 : タグセットを集中管理し各組織にタグの追加を認めない

レベル 2 : タグセットの一部を集中管理する

レベル 3 : 各組織が全く自由にタグセットを決められる

レベル 1 では、タグを利用した情報収集が容易であるが、組織独自のタグを利用できないため、多様な視点からの情報を共有するには向きである。レベル 3 では、各組織は独自の視点からの情報を記述するためにタグを利用できるが、他の組織が情報を収集するだけでもタグの変換が要求される。レベル 2 では、集中管理されるタグを利用した情報収集ができる、かつ独自情報の記述も可能である。レベル 2 はレベル 1 とレベル 3 の利点を併せ持つため、多様な視点からの情報を記述した XML データの共有に有効であると思われる。そのため本研究ではレベル 2 のアプローチを採用した。

2.2 運用方針

本研究では、共有される XML データについて、上位のタグセットを集中管理する。これをテンプレートとし、その下で各組織は自由にタグを追加することができる。

まず、情報共有に係わる組織は協力してテンプレートを作成する。各組織はテンプレートを構成するタグセットに組織独自のタグを追加し、追加したタグの一覧(タグセットとどのような情報を記述するか)を他の組織に報告し、自身の情報を検索する機能を公開する。各組織は報告されたタグの一覧を見て、どの情報を利用するかを決定する。

2.3 アーキテクチャ

提案するアーキテクチャは、アプリケーションサーバ、タグ管理サーバ、データサーバにより構成され、各サーバ間は SOAP(Simple Object Access Protocol)を利用してデータをやり取りする(図 1)。SOAP は XML でエンコードされたデータを HTTP(HyperText Transfer Protocol)や SMTP(Simple Mail Transfer Protocol)などインターネット標準プロトコルでやり取りするためのプロトコルであり、元々はインターネット上で RPC(Remote Procedure Call)を実現するために開発された[11]。

・アプリケーションサーバ

データサーバのデータを利用する組織独自のアプリケーションシステムである。各組織が登録したタグの一覧をタグ管理サーバから受け取り、データサーバの検索機能を呼び出す。データサーバから受け取ったデータをどのように処理するかはアプリケーションに依存する。

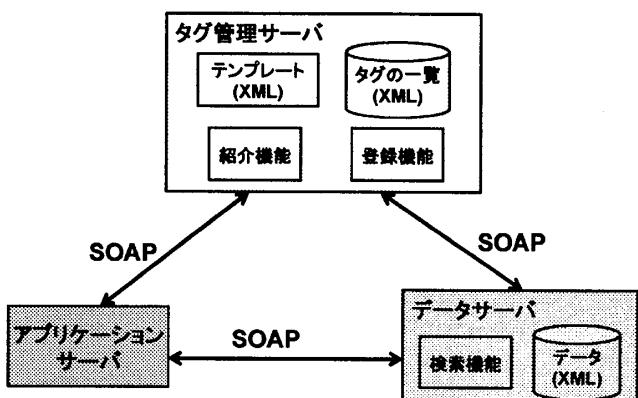
† 神奈川工科大学 情報学部 情報工学科

・タグ管理サーバ

XML によりテンプレートと報告されたタグの一覧、およびタグの一覧ごとにデータサーバ上で検索機能を提供する SOAP プログラムのアドレスを管理し、それらの情報をアプリケーションサーバに提供する(紹介機能)。また、組織の担当者がその組織で追加したタグの一覧を登録するための機能を持つ。紹介機能と登録機能はテンプレートを利用する。これらの機能により、各組織は他の組織にテンプレートの下に追加したタグの一覧を報告することができる。

・データサーバ

何らかの情報を収集している組織によって管理されるサーバであり、テンプレートを構成するタグに基づきデータを検索する機能を提供する。この機能を実装する SOAP プログラムのメソッド名をサーバ間で統一にする。検索結果は、各組織で管理する XML の形式で返される。データサーバを管理する組織によってデータサーバがアプリケーションサーバになることもある。



3. 福祉情報を利用したシステムの試作

3.1 福祉情報提供システム

近年、入口付近に段差があるかどうかなど、店舗や公共施設のバリア・バリアフリー情報(以下、福祉情報)を記載した福祉マップが数多く作成され、Web 上で公開されているものも存在する。福祉情報は障害者や高齢者のみでなく多くの人に有用である。現在、様々な組織あるいは個人で情報を収集しているが、一組織や一個人での情報収集には限界がある。そのため、これらの情報を共有可能にすれば、そのような問題の解決につながる。また、様々な組織が独自の視点で福祉情報をを集めているが、いくつかの項目は共通している。

これらの理由から、システムの試作では福祉情報を対象にした。システムの試作には.NET Framework を利用し[12]、施設名や住所で福祉情報を検索するアプリケーションを作成した。

3.2 タグ管理サーバの構造

(1) テンプレートとタグの一覧の構造

福祉情報を収集しているいくつかのボランティア組織がそれぞれに利用している福祉情報の調査票を参考にし、テンプレートのタグ構造を仮想的に図 2 のようにした。図 2 でアスタリスク(*)が付いているタグは、組織がその下に新たにタグを追加できないということを示し、各データサーバはアスタリスクが付いたタグを指定した検索機能を SOAP により提供する。

一方、各組織でどのようなタグを追加したかを管理するため、テンプレートの facility 要素の子要素のタグから、追加されたタグまでのツリー構造をそのまま利用する。たとえば、ある組織が図 3 に示した構造で福祉情報を蓄積(提供)しているとすると、その組織のタグの一覧は図 4 のように記述される。図 2 と図 4 から、この組織は toilet 要素の下に新たに三つのタグを追加し、そのうちの一つはさらにその下にタグを持っているということが分かる。

(2) 登録機能

タグ管理サーバの登録機能は、組織が追加したタグを抽出する部分、ユーザインターフェースを生成する部分、タグの一覧を XML 形式で追加する部分からなる Web アプリケーションである。

組織の担当者がデータサーバ上の SOAP プログラムのア

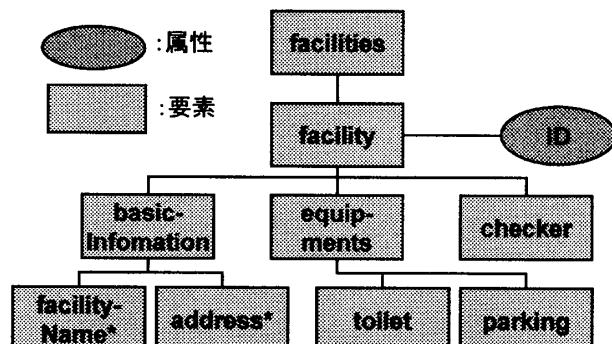


図2 テンプレートのタグ構造

```

<facility id="1">
  <basicInformation>
    <facilityName>市役所</facilityName>
    <address>〇〇市△△××丁目</address>
  </basicInformation>
  <equipments>
    <toilet>
      <location>玄関を取り市民課の横を通り抜けた所</location>
      <type>多目的トイレ</type>
      <others>
        <comment>きれいに清掃されている</comment>
      </others>
    </toilet>
    <parking>あり(3台)</parking>
  </equipments>
  <checker>福祉まちづくりの会</checker>
</facility>

```

図3 福祉情報の例

```

<dataSource ID="識別番号" date="登録日"
    URL="SOAPプログラムのアドレス">
<equipments>
<toilet>
<location>施設内のトイレの場所</location>
<type>トイレの形状</type>
<others>
<comment>使用者のコメント</comment>
</others>
</toilet>
</equipments>
</dataSource>

```

図4 図3に対応するタグの一覧

ドレスを入力すると、追加タグ抽出部は SOAP プログラムのメソッドを呼び出して検索を実行し、検索結果から facility 要素を一つ取り出し、テンプレートと比較することでその組織で追加されたタグを抽出する。抽出結果は、インターフェース生成部を介して、そのタグでどのような情報を記述しているかを入力するためのテキストボックスとそのタグで記述された情報を共有するかどうかを指定するためのチェックボックスとともに表示される。組織の担当者がタグで記述される情報の内容を入力すると、タグ一覧追加部は入力内容から図 4 のような XML 形式の要素を生成しタグの一覧に追加する。

(3) 紹介機能

タグ管理サーバは各組織のタグの一覧を提供する機能を持つが、アプリケーションサーバは SOAP によりこの機能を利用することができる。アプリケーションサーバがタグ管理サーバにタグの一覧を要求すると、要求あるいは登録された数だけ図 4 に示した構造のタグの一覧をグループ化した XML 要素が返される。この情報を利用することでアプリケーションサーバは追加されたタグの一覧を利用者に表示したり、各データサーバ上の SOAP プログラムのメソッドを呼び出すことができる。

3.3 試作システムの実行例

図 5 は、タグの一覧の登録画面において図 3 に示した構造情報を提供する組織の SOAP プログラムのアドレスを入力した結果であり、location や type などその組織が追加したタグが抽出されている。そのため、組織の担当者は追加したタグの中から共有可能な情報を記述するタグのみの一覧を容易に登録することができる。

図 6 はアプリケーションサーバがタグの一覧を利用してデータサーバの検索機能を呼び出し、検索を実行した結果である。図 6 では、検索対象となったデータサーバを管理する組織がテンプレートの下に追加したタグによって記述された情報の部分を点線の枠で示した。このように組織独自のタグで記述された情報も共有可能である。

4. 考察

(1) テンプレートの導入について

テンプレートを導入したことにより、各組織はテンプレートの下に自由にタグを追加することができ、また、タグ管理サーバを通じて相互に追加したタグを報告することで、

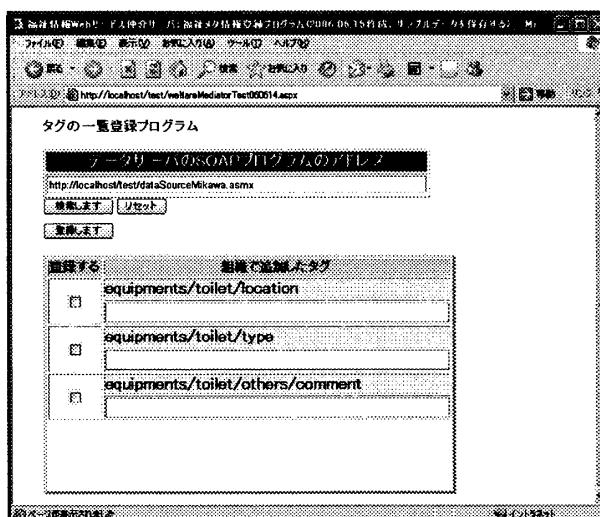


図5 タグの一覧の登録画面

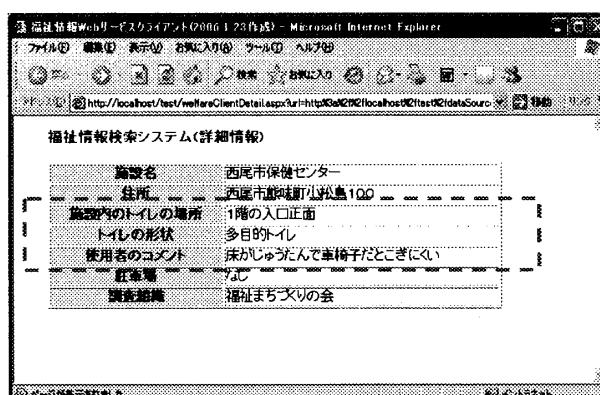


図6 福祉情報の検索例

そのタグで記述されたデータを容易に交換することができた。従来の研究では、共通のタグセットで記述することができないデータを交換することは困難であったが、本システムはその問題を解決できる。

さらに、各組織の視点で集めた情報をテンプレートの下にまとめ合わせるといったことも期待できる。しかしそのためには、各組織の情報が同じものを対象にしていることが分かるようにする必要があり、テンプレートのタグセットだけでなくその内容についても集中管理するなどの工夫が必要である。

(2) タグの一覧の管理方法について

本研究では、情報共有に係わる組織はどのようなタグを追加したかをタグ管理サーバに登録することで他の組織に報告するというアプローチを探っている。各組織が DTD(Document Type Definition)を管理するという方法も考えられるが、この方法では、タグが変更された場合、他の組織がそのことを知ることが困難である。

一方、タグ管理サーバではテンプレートの下にどのようなタグが追加されたかを管理する。そのため、タグの一覧が登録された後でも定期的にデータサーバから XML データを取得し、タグ管理サーバ上の情報を更新できる利点がある。共通のタグセットやオントロジーとの対応関係を管理する場合、その対応関係を適切に維持することは困難である。

また本研究では、情報共有に係わるすべての組織からタグの一覧を見ることができる。タグの一覧を登録する際にその情報を見ることができる組織を指定し、特定の組織とのみ情報を共有するという方法も考えられる。しかし多様な視点からの情報を集めるという点では、すべての組織からタグの一覧を見ることができたほうが有利である。

(3) タグの一覧の登録について

試作システムでは、タグの一覧を登録する際、タグ管理サーバが入力された SOAP プログラムのアドレスを利用してデータサーバの検索機能を呼び出すことで、組織で追加されたタグを抽出することができた。したがって、組織の担当者は容易にタグの一覧を登録することができる。また、タグを抽出する仕組みは、データサーバ上でタグの変更がないかどうかを確認し、タグの一覧を最新の状態に保つことにも応用できる。しかし試作システムでは属性や、文字データと要素を両方含む混在要素などを考慮していない。多様な視点からの情報を共有するためには、これらを含んだ XML データの場合でもタグの一覧を登録できるようにする必要がある。

(4) 各組織のサーバ上での検索について

提案したアプローチでは XML データの検索は、各組織が管理するサーバ上で行われるが、タグ管理サーバが各組織から XML データを集めて検索するという方法も考えられる。しかしこの方法では、他の組織と共有したくない情報もタグ管理サーバに集められたり、それを避けるためには共有する情報と共有しない情報を組織の側で分けて管理する必要があり手間がかかるなどの問題がある。

また本研究では SOAP により検索機能を提供するようにしている。そのため、アプリケーションプログラムの開発言語やオペレーティングシステムに依存することなく検索機能を呼び出すことができるため、様々な組織との情報共有に適している。

5. まとめと今後の課題

本研究では、各組織で分散管理されている XML データを共有するためのテンプレートベースのアーキテクチャを提案した。このアーキテクチャは、各組織の XML のツリーコンストラクチャについて、上位のタグセットを集中管理し、これをテンプレートとして、それより下位のタグについては各組織が自由に追加できるというアプローチに基づいている。そのため、共通のタグを利用する場合に比べて多様な視点からの情報を組織間で共有することができる。

提案したアーキテクチャはアプリケーションサーバ、タグ管理サーバ、データサーバから構成され、それらの間では SOAP により XML データがやり取りされる。本研究では福祉情報を対象としてシステムを試作した。試作システ

ムでは、組織で追加されたタグをタグ管理サーバに登録することが容易に行え、また、そのタグで記述されたデータを組織間で交換することができた。

しかし、多様な視点からの情報をまとめ合わせるには、情報の対象が同じであることが分かるようにするためにテンプレートのタグの内容も決めたり、属性や混在要素を含む XML データを処理したりするなどの問題があり、これらを解決することが今後の課題である。また、本研究で採用したアプローチでは、情報共有に係わる組織はすべての組織に追加したタグを報告するため、すべての組織に対等に情報を提供することになる。しかし実際には、組織間での交渉を通じて共有する情報を決めるという状況も考えられる。そのため、今後はこのようなアプローチに基づくシステムも検討する必要がある。

参考文献

- [1] Anant Jhingran, Nelson Mattos, and Hamid Pirahesh: Information integration: A research agenda, IBM Systems Journal, Vol 41, No. 4, pp.555-562 (2002).
- [2] Elisa Bertino, Elena Ferrari: XML and Data Integration, IEEE Internet Computing, Vol.5, No.6, pp.75-76 (2001).
- [3] Sophie Cluet, Pierangelo Veltri, Dan Vodislav: Views in a Large Scale XML Repository, Proceedings of the 27th VLDB Conference, pp.271-280 (2001).
- [4] Young-Kwang Nam, Joseph Goguen, Guilian Wang: A Metadata Integration Assistant Generator for Heterogeneous Distributed Databases, Proceedings of Confederated International Conferences CoopIS, DOA, and ODBASE 2002, pp. 1332-1344 (2002).
- [5] Bernd Amann, Catriel Beeri, Irini Fundulaki, and Michel Scholl: Ontology-Based Integration of XML Web Resources, International Semantic Web Conference 2002, pp.117-131 (2002).
- [6] Laks V. S. Lakshmanan, Fereidoon Sadri: XML Interoperability, WebDB 2003, pp.19-24 (2003).
- [7] Serge Abiteboul, Omar Benjelloun, Tova Milo: Web Services and data integration, WISE 2002 (2002).
- [8] Mark Hansen, Stuart E. Madnick, Michael Siegel: Data Integration using Web Services, DiWeb 2002, pp.3-16 (2002).
- [9] Daniela Leal Musa, Jose Palazzo de Oliveira: Sharing Learner Information through a Web services based, Web Information Systems Modeling (WISM) 2004, pp.122-131 (2004).
- [10] Hyun Kim, Hyung-Sun Kim, Joo-Haeng Lee, Jin-Mi Jung, Jae Yeol Lee and Nam-Chul Do: A framework for sharing product information across enterprises, The International Journal of Advanced Manufacturing Technology, Vol.27, No.5-6, pp.610-618 (2006).
- [11] Thomas Erl: Service-Oriented Architecture, PRENTICE HALL (2005).
- [12] 矢島聰: VB.NET による XML Web サービス開発, インプレス (2003).