

Title	Efficient XML Data Processing in Web Services(Abstract_要旨)
Author(s)	Takase, Toshiro
Citation	Kyoto University (京都大学)
Issue Date	2008-09-24
URL	http://hdl.handle.net/2433/123822
Right	
Type	Thesis or Dissertation
Textversion	none

(論文内容の要旨)

本論分では、WebサービスにおけるXMLデータ処理の効率化について論じる。Webサービスとは、ソフトウェアを再利用に適した機能単位でコンポーネント化し、これらのコンポーネントを遠隔サイトからもアクセス可能なようにネットワーク上のサービスとしてインターネット上に公開したものである。

ソフトウェアをWebサービス化する目的は、ソフトウェアの再利用性を従来以上に高めることであり、よって、ソフトウェアをWebサービス化するにあたっては、不特定のユーザが、様々な環境、プラットフォームからアクセス可能であるようにすること、すなわち、高い相互運用性を実現することが重要である。そのため、現在、Webサービスのための様々な標準規格が策定されており、これらの標準規格に従った形でWebサービスを構築、公開されている。

これらの標準規格では、標準データ形式としてXMLが採用されている。

これは、XMLが可読性のあるテキスト形式を採用しており、かつ、様々な文字コードに対応しているため、特定のプラットフォームに依存しない高い相互運用性や、様々なWebサービスを実現する上での高い開発効率を実現する上で適しているためである。

しかし、データ形式としてXMLを用いたWebサービスは、これまでに提案されてきた他の様々な分散計算の枠組みと比べて、処理効率が劣ることが指摘されている。これは、XMLが上述のように様々な文字コードを含みうる可読なテキスト形式であるため、各Webサービス内部の処理においては、これを逐一、そのWebサービスが実装されているプラットフォーム上のデータ形式に変換する必要があるため、このコストが非常に高いためである。

しかし、これは高い相互運用性を実現するために特定のプラットフォームに依存しないデータ形式を採用する場合には不可避なコストである。

現状のWebサービスに関する以上の観察から、本論文ではWebサービスにおけるXMLデータのデータ変換処理のコストを低減する技術について論じる。

そのような技術を開発する上で重要な鍵となるのは、WebサービスにおけるXMLデータ処理は同じ処理、あるいは冗長な処理の繰り返しが非常に多いという点である。この点に着目し、本論分では大きく以下の二つのアプローチによって、XMLデータの処理コストを軽減する。

- (1) 同じ処理に対して過去の処理結果を再利用する。
- (2) 予め冗長な処理が起きないように工夫をほどこす。

前者のような、過去の結果を再利用する手法としては、様々なキャッシュの手法がこれまでに研究されている。しかし、従来のキャッシュ手法を単純にWebサービスに対して適用するのでは、必ずしも高い処理効率が得られない。これは、WebサービスにおけるXML処理では、データの変換が何段階かに分けて行われ、過去の結果をキャッシュする際には、これらのうちのどの段階の結果をキャッシュするかによって、全体の処理効率が大きく異なるためである。

本論分では、各データの特性とWebサービスのセマンティクスを考慮して、自動的に適切な段階でキャッシュを行う手法について提案する。

また、Webサービスにおいては、非常に良く似た、しかし少しだけ違うようなXMLデータのパース処理を行うことが多い。これは、Webサービスのサーバに多くのアクセスが集中するような状況は、同じサービスに対する、一部の引数だけが異なるような多数の呼び出しが集中することによって生じることが多いためである。

そこで、本論分ではXMLデータをその木構造に基づいて分割し、各部分毎に過去のパース結果をキャッシュすることによって、完全に同じではないが共通部分が多いようなXMLデータの処理を効率化する手法について提案する。

前述の二つのアプローチのうち、後者の冗長な処理の回避に関しては、まず同じサービスに対する複数の呼び出しを一つの呼び出しにまとめることで効率化する手法を提案する。Webサービスにおいては、サービスとして提供する処理の単位を細かくするほど、データ変換処理などのオーバーヘッドが高くなり、処理の単位を大きくするほどオーバーヘッドが低くなるという性質がある。

そのため、処理効率だけを考えれば、処理の単位をできるだけ大きくすべきであるが、ソフトウェアコンポーネントの再利用という本来のWebサービスの目的を考えれば、処理単位は再利用に適した論理的な処理単位に基づいて決められるべきである。そこで、本論分では、サービスの実装が提供するインターフェイス、およびそれを呼び出すユーザに見せるインターフェイスのレベルでは処理の単位を変更せずに、これら二つの各々にラッパーをかぶせることによって、実際のサービスの呼び出し処理においては、複数の呼び出しを一つにまとめて効率化する手法を提案する。

また、Webサービスにおいて頻繁に生じる冗長な処理のもう一つの例として、ネットワークから受け取ったあるXMLデータのパースと、そのデータを再度、ネットワークに送り出すためのXMLデータへの再変換という、逆変換の繰り返し処理がある。そこで、本論分では、XMLデータとそのパース結果を対にして保持することにより、このような逆変換を回避する手法を提案する。

この手法では、XMLデータのうち実際に必要な部分のみをパースすること、また、パース結果に対する更新操作のXMLデータへの伝播を最終的なXMLデータへの逆変換まで遅延することにより、高い処理効率を実現する。

本論分では、以上に説明した四つの手法により、Webサービスにおいて問題となる、XMLデータの変換処理部分の高いコストを低減することを提案し、これらの手法により、実際に処理効率が大きく改善されることを実験によって示す。

氏名	高瀬 俊郎
----	-------

(論文審査の結果の要旨)

本論分はWebサービスの処理の効率化に関するものである。Webサービスは、ソフトウェアを各ユーザの計算機にインストールして利用するという従来の計算機の利用形態を、各ユーザはネットワーク上のサービスを遠隔呼び出しにより呼び出すことによって計算機を利用するという利用形態へ転換するという、情報技術の根幹に関わる重要な研究分野である。

また、あらゆる機器がネットワークで接続され、携帯型情報端末を持ったユーザが社会のあらゆる場面で、これらのネットワークを通して情報にアクセスするという、今後のユビキタス社会の実現に向けて、社会的にも非常に大きな影響を持つ技術となっている。このようなWebサービス技術の実社会での進展において、その処理効率の改善は、現在、重要な問題となっており、よって、本論分で論じられている問題は、技術的な見地からも社会的な見地からも、重要な問題であると言える。

本論分では、まず、Webサービスの処理効率が低い主たる原因が、Webサービスの相互運用性を確保するためには不可避な、プラットフォーム非依存の標準データ形式と各プラットフォーム上のデータ形式との間の変換コストにあることを示し、また、Webサービスにおいては同じようなデータ変換の処理の繰り返しが非常に多いこと、よって、この部分の冗長な処理のコストを省くことによって、Webサービスの処理効率を高めることは原理的に可能と考えられることを指摘している。

この考え方に基づき、実際のWebサービスに頻繁に現れる冗長な処理のコストを軽減する四つの手法について提案し、それらの手法が実際にWebサービスの処理コストを低減することを実験によって示している。

以上から、本論分は、Webサービスの効率的な処理という技術上および社会上の観点から重要な問題に関して、今後の研究開発に対して貢献しうる有効な知見を示すとともに、Webサービス技術の実用において広く利用可能な具体的な処理手法を示しており、技術上、社会上の両面において有用な研究成果を示していると言える。

また、本論分で示されている四つの手法については、それぞれが、非常に重要度の高い国際会議を含む四つの国際会議において採択、発表されており、この分野の国際的な学術コミュニティにおいても高く評価されていると言える。

よって、本論文は博士(情報学)の学位論文として価値あるものと認める。

また、平成20年8月28日実施した論文内容とそれに関連した試問の結果合格と認めた。