

仕様の異なるXML文書間の変換言語 ～変換の「正しさ」の理論による保証～



上級准教授 鈴木 大郎

概要

○XML文書変換の正しさの検証

インターネットの世界では、ある仕様に従うXML文書から別の仕様に従うXML文書への変換が頻繁に行われます。そのような変換を記述する言語としてXSLTが有名です。しかし、XSLTで記述された変換プログラムは、仕様に関係なく変換を行うので、仕様を満たさないXML文書を生成してしまう場合があります。また、仕様は満たしていても望ましくない文書を生成したり、変換そのものが停止しないこともあります。

変換プログラムがこのようなことを起こさない、つまりプログラムが行う変換が「正しい」ことを、変換を行う前に検証できるようなXML文書変換言語を設計しています。

○理論に基づいた言語設計

変換プログラムの正しさを保証するためには、変換の振る舞いを解析するための理論が必要になります。

ここでは、「型理論」と「項書き換え系」と呼ばれる理論をベースとして、その理論を構築し、それに基づいて言語を設計しています。また、理論に基づいて言語を設計することで、変換の正しさを保証しつつ、従来の変換言語では記述しづらいような強力な変換方法を提案することができます。

実用化の可能性

○高階項書き換え系と型理論に基づくXML文書変換言語の最初の版の設計が終了し、それに基づく言語処理系のプロトタイプが構築されています。

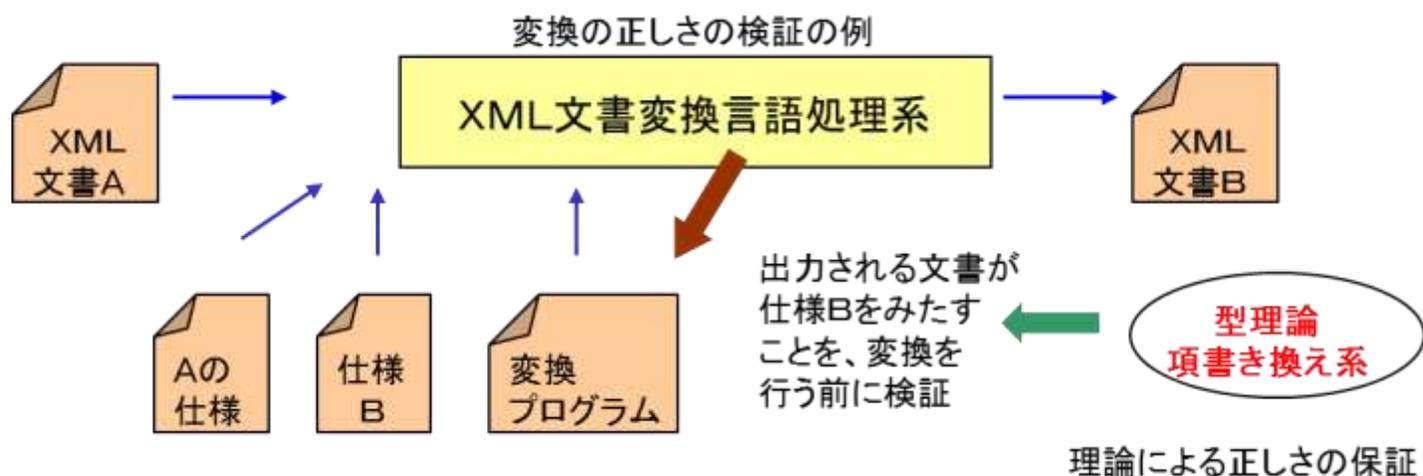
○いまのところ、変換プログラムが生成するXML文書が仕様を満たすことの検証が可能で（研究概要図を参照）。今後、変換が停止することの検証などを行っていきます。

○高階項書き換え系を利用することで、現在の版でも従来の変換言語よりも強力な変換が可能になっています（文脈の抽出を用いた変換）。現在、より強力な変換を可能にするよう、理論に改良を加えています。

UBICからのメッセージ

○ここで言う「変換」は一般のWEBユーザには見えない所で行われています。WEBページはその内容を記述するのに便利なXML言語で書かれています。この種の文書は普通のパソコンにとって、多くの場合、非常に扱いにくいものです。その文書を「変換」することにより初めてコンピュータが扱いやすい文書になります。その変換言語が鈴木上級准教授の研究テーマの一つです。この研究成果はWEBサービスを提供している企業が必要としています。

研究概要図



正しく強力なXML文書の変換を実現