

# 要 旨

## XPath クエリと木パターン照合問題の 関係に関する研究

前田 秀樹

XML は、木構造をなすデータを表現するマークアップ言語であり、データ交換やデータベースのためのフォーマットとして広く使われている。XML 文書の要素を指定するために、パス式に基づく言語 XPath が広く用いられている。例えば、XML というタイトルの本を指定するのに “//book[title='XML']” という XPath 式を用いることができる。

木パターン照合問題は、木構造データに対する重要な問題の 1 つである。これは、パターン木とターゲット木の 2 つが与えられたとき、ターゲット木の中でパターン木が含まれている部分を全て探す問題である。木パターン照合問題に対して、これまでに順序木として解くアルゴリズムと無順序木として解くアルゴリズムが示されている。

XPath クエリと木パターン照合問題は、いずれも木構造からノードを選択するものである。本研究の目的は、XPath クエリのうち木パターン照合問題によって計算できるものを示すことである。そのために、XPath 式からパターン木を生成するアルゴリズムを示す。実用によく用いられる XPath 式において、兄弟間の順序を考慮する必要のある部分は式の一部であることが多い。そこで、本研究では、兄弟間の順序の考慮をノードごとに指定したパターン木を定義する。この定義によって、XPath 式から生成されるパターン木の個数を減らすことができる。

本研究で示したアルゴリズムを用いて、XPath 式からパターン木が生成されることを確かめることにより、木パターン照合問題によって計算できる XPath クエリの内容を示した。また、提案アルゴリズムを評価したところ、XPath 式によってはパターン木が必要以上に

多く生成されることが分かった。この理由は、提案アルゴリズムにおいて XPath 式に兄弟軸が含まれていることを前提とした処理を行っているためであると考えられる。

キーワード XML, XPath, 木パターン照合問題

# Abstract

## A Study of the Relationship Between XPath Query and Tree Pattern Matching Problem

Hideki Maeda

XML a markup language for representing tree structured data, is widely used for the format of data exchange and databases. To specify the elements in an XML document pass-expression-based language called XPath is widely used. For example, the XPath expression “`//book[title='XML']`” specifies the book whose title is XML.

The tree pattern matching problem is an important problem for tree structured data. Given two trees, the pattern tree and the target tree, this problem is to find all the parts in the target tree that contain the pattern tree. So far, algorithms for ordered trees and unordered trees have been developed.

Both XPath queries and the tree pattern matching problem are to select nodes in a tree. The goal of this study is to show a certain set of XPath queries can be computed as the tree pattern matching problem. I show an algorithm that generates the pattern tree from XPath expressions. Usually only parts of XPath expressions specify the order between sibling. Therefore, in this study, I give another definition of the pattern trees in which we may specify the order between sibling for some nodes.

I showed a set of requirements of XPath queries so that we can compute then as the tree pattern matching problem by confirming that the proposed algorithm surely generates patterns from XPath queries. The algorithm may generates wore pattern trees than necessary. This is due to the assumption that all the input XPath expressions have

intra-sibling axes.

*key words* XML, XPath, Tree pattern matching problem