

発表概要

テキスト値出現位置に着目した XML 文書のスキーマ変換器学習手法

小川 順平^{1,†1} 粕谷 英人^{2,a)} 大久保 弘崇² 山本 晋一郎²

2012年1月23日発表

XML スキーマ変換器を自動生成する手法を提案する。ここでいう変換器とは、あるスキーマ A に基づく XML 文書を入力とし、同じ情報を表す別のスキーマ B に基づく XML 文書を出力するプログラムである。提案手法は、スキーマ A と B に基づく同一内容の XML 文書対を入力として変換器を生成する。提案手法は、両文書中の等しいテキスト値の出現位置を利用して対応付けを行い、スキーマ A の木構造から取り出したテキスト値をスキーマ B の木構造に埋め込む変換を生成する。提案手法が必要とする入力は一つの文書対と文書それぞれのスキーマである。スキーマからは構造の横方向の繰返しを判断する。既存研究では XML 要素の名前の類似性や構造の類似性を用いたものが多く、これらはスキーマを部分的に改変した場合などを対象にしている。これに対して、提案手法は構造が大きく異なるスキーマ間や A と B で同一の XML 要素名が異なる意味で用いられている場合でも変換器を生成できる。

Program Synthesis of XML Schema Transducer from Instances Using Positions of Text Values

JUNPEI OGAWA^{1,†1} HIDETO KASUYA^{2,a)} HIROTAKA OHKUBO²
SHINICHIRO YAMAMOTO²

Presented: January 23, 2012

This presentation proposes a method to automatically generate XML schema transducer. A transducer is a program that takes an XML document based on some schema A as input and outputs an XML document based on another schema B while preserving information. Our method generates a transducer by a pair of XML documents that contains the same information based on schema A and B respectively. First, our method makes a mapping by using position of each occurrence of a text value included in both documents. Next, it generates a transducer that extracts text values from tree structure of schema A and embeds them into the tree structure of schema B. Moreover, our method is capable of handling recurrence in schema structure. Most of related works that transduce schemas use similarity of names of XML elements or similarity of schema structure. Their target is partly modified schema. Whereas, our method can generate transducer, even in the case of greatly different schemas or if same names of XML elements are used for different meaning in schema A and B.

¹ 愛知県立大学大学院情報科学研究科
Graduate School of Information Science and Technology,
Aichi Prefectural University, Nagakute, Aichi 480-1198,
Japan

² 愛知県立大学情報科学部
School of Information Science and Technology, Aichi Prefec-
tural University, Nagakute, Aichi 480-1198, Japan

^{†1} 現在, KVH 株式会社
Presently with KVH Co., Ltd.

^{a)} kasuya@ist.aichi-pu.ac.jp