

XQuery Fusion; XMLストアの静的エミュレーションに基づくXQuery融合変換

An XQuery Fusion based on Static Emulation of XML Store

加藤弘之 日高宗一郎 胡振江 中野圭介(電通大) 石原靖哲(阪大)
Hiroyuki KATO Soichiro HIDAKA Zhenjiang HU Keisuke NAKANO Yasunori ISHIHARA

何ができる？

XMLはデータ交換フォーマットです。XQueryはXMLデータに対する関数型の問合せ言語です。本研究では、入力XQueryの式を、効率よく実行できるXQueryの式に書き換えることを研究しています。これにより、XMLによるデータ統合が効率的に実現できます。

現状と課題

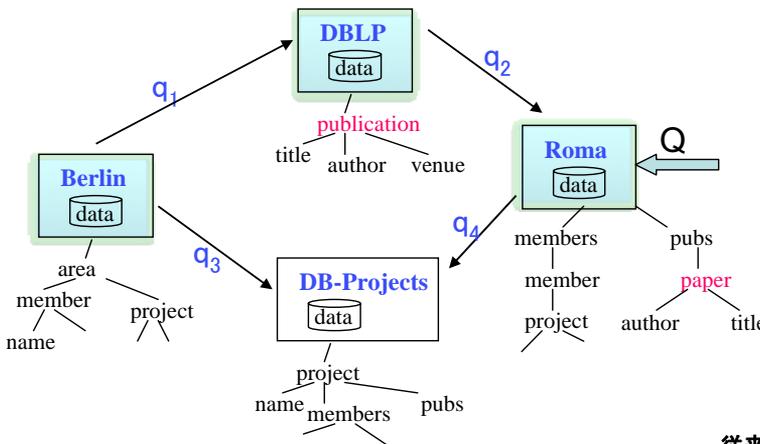
現状:

- 書き換えアルゴリズムの開発
- プロトタイプ実装
- 人工データを用いた予備実験による定量評価

課題:

アルゴリズムの正しさ, 有効範囲などを示す。

XMLによるデータ統合



q_1, q_2, q_3, q_4 は, XQueryで記述されたスキーママッピング
例えば q_2 は, DBLPのデータをRomaに変換するXQuery式
(データベースにおけるビュー定義に相当) (Global-As-View)

Romaに対する問合せQは, スキーママッピングで繋がっている他のサイトが管理しているデータを検索可能. (P2Pデータ統合)

問合せQの結果 := $Q(\text{Roma}) \cup Q(q_2(\text{DBLP})) \cup Q(q_2(q_1(\text{Berlin}))) \cup \dots$

$Q(q_2(\text{DBLP})) = Q \circ q_2(\text{DBLP})$ 一般に $Q \circ q_i$ には冗長な式が存在

$Q: /\text{pubs}/\text{paper}/\text{author}$ とすると,

RomaにあるpaperがDBLPにはないため, Qは以下の式を含む.

(1): $\langle \text{paper} \rangle (\$v/\text{author}, \$v/\text{title}) \langle / \text{paper} \rangle / \text{author}$

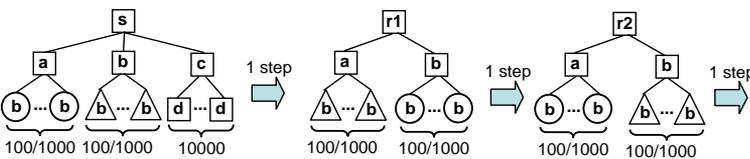
従来の融合変換を適用し, 冗長な式を取り除くと, (2): $\$v/\text{author}$

(1)と(2)はXMLの特徴的な意味であるノードIDに基づいた等価性を保持していない。本研究ではノードIDの等価性を維持した融合変換が目標。

問合せQを記述する際, Romaのスキーマだけを知っていれば良い

予備実験結果

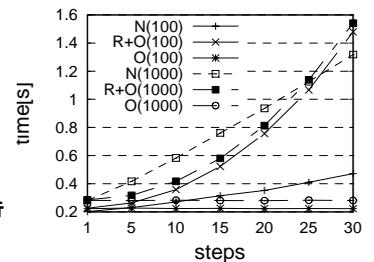
実験1:



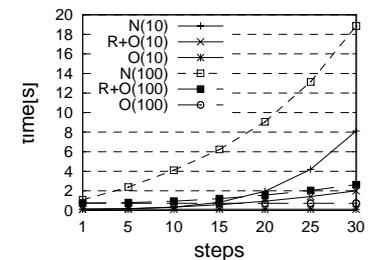
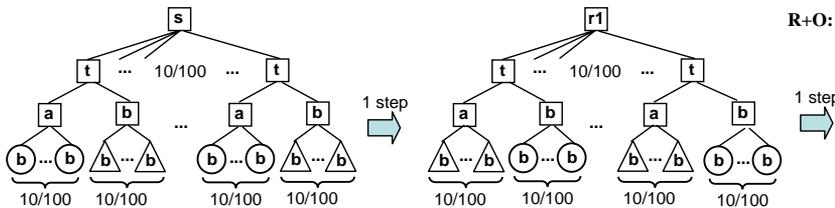
N: 入力問合せをそのまま実行

O: 入力問合せを書き換えて実行 (書き換え時間は含まず)

R+O: 入力問合せを書き換えて実行 (書き換え時間を含む)



実験2:



実験1の結果より, ステップ数が増えたとき(25ステップ)に書き換えコストが増大し, 入力問合せをそのまま実行した方が速くなるのがわかった。現在アルゴリズム修正中。

発表文献及び研究予算

H. Kato, S. Hidaka, Z. Hu, Y. Ishihara, K. Nakano, Rewriting XQuery to Avoid Redundant Expressions based on Static Emulation of XML Store, ACM SIGPLAN Workshop on Programming Language Techniques for XML (PLAN-X 2009), Savannah, Georgia, USA, Jan. 24, 2009.

本研究の一部は, 文部科学省科学研究費補助金(基盤研究(C) 課題番号20500043「XMLデータ統合問題解決のためのXQueryの静的解析に基づく書き換え手法の開発」研究代表者:加藤弘之)によるものです。