

氏名	池田 光雪				
学位の種類	博士（情報学）				
学位記番号	博 甲 第 7918 号				
学位授与年月日	平成 28 年 7 月 25 日				
学位授与の要件	学位規則第 4 条第 1 項該当				
審査研究科	図書館情報メディア研究科				
学位論文題目	A Study on Algorithms for Finding Correct XPath Queries（XML 問合せ式における修正候補発見アルゴリズムに関する研究）				
主査	筑波大学	教授	博士（工学）	佐藤哲司	
副査	筑波大学	准教授	博士（工学）	鈴木伸崇	
副査	筑波大学	教授	博士（工学）	森嶋厚行	
副査	筑波大学	教授	工学博士	杉本重雄	
副査	筑波大学	准教授	博士（工学）	天笠俊之	

## 論文の要旨 (2,000 字程度)

データベースは、大量のデータを効率的・永続的に管理し、問合せによって所望の結果を得ることができる。所望の結果を得るための正しい問合せを記述することは、一般的なデータベース利用者には困難が伴うことから、適切な問合せの記述を支援する手法が研究されている。関係データベースでは、標準化された問合せ記述言語 SQL の記述を支援する、数多くの手法が提案され、実用に供されている。一方、構造化文書を対象とする XML データベースでは、問い合わせ言語 XPath は標準化されているが、問合せ対象が複雑な構造を有していることなどから、適切な XPath 式を記述するための支援手法は研究段階にあるといえる。

問合せ式の記述を支援するには、対象となる XML 文書群のデータ構造を正確に把握する必要がある。そのような情報を得る方法として、XML 文書群を走査して情報を収集する方法と、スキーマを参照する方法の 2 つが考えられる。一般に、大量のデータに対する走査は、多くの計算機資源を必要とすることから、前者の手法は、計算資源の限られた環境では現実的な時間で処理できないなどの問題がある。本論文では後者のスキーマを参照する手法に着目し、スキーマに基づく問合せ式の記述支援、すなわち、与えられたスキーマの下でユーザが入力した XPath 式を修正する問題(XPath 式修正問題)について考察している。スキーマに基づいて問合せ式の振る舞いを解析することを静的解析といい、XML における静的解析は多くの場合において計算困難であることが知られている。本論文で考察する XPath 式修正問題についても、この問題が効率よく解けるか否かは自明でなく、アルゴリズム・計算複雑性の観点からこの問題の性質を究明することは重要な課題である。そこで本論文は、XPath 式修正問題において、計算困難な部分と効率よく解ける部分を明らかにすることを主たる目的としている。

XPath 式に対する編集操作と XPath のクラスは、XPath 修正問題の計算複雑さに影響を与える主要な要因である。本論文では、第 4 章において XPath に対する 6 種の編集操作：(1)軸の置換、(2)ラベルの置換、(3)ロケーションステップの挿入、(4)ロケーションステップの削除、(5)ロケーションステップの交換、(6)述語の交換、を導入している。ここで、編集の基本操作である(1)から(4)までを core といい、(5)および(6)を extended という。また、XPath 式の部分クラスとして、simple および XP を定義する。XP は軸として child、descendant-or-self、following-sibling、preceding-sibling、attribute を用いた XPath 式の集合である。また、XP に属しかつ述語も attribute 軸も使用しないものを simple な XPath 式と呼ぶ。これら定義に基づいて、XPath 式修正問題を「DTD の下で入力 XPath 式に最も近い(編集距離の最も小さい)XPath 式を求める問題」と形式化し、この問題の計算複雑さを明らかにしている。まず、編集操作として core および extended を許す場合、XPath 式を simple なものに限定しても本問題が NP 困難であることを示している。次に、ロケーションステップの交換を用いないと仮定しても、述語の交換を許した場合、ある XP の部分クラスに対して本問題が NP 困難となることを示している。一方、編集操作を core に限定した場合、XP に属する XPath 式に対して、多くの場合において本問題が効率良く解けることを後述のアルゴリズムにより明らかにしている。

XPath 式の修正を行う場合、その式に近くかつ構文的に正しい XPath 式は複数存在し得るため、そのような XPath 式の中から望ましい式を選択できることが望ましい。そこで本論文では、編集操作を core に限定した上で、DTD の下で入力 XPath 式に最も近い K 個の XPath 式を求めるアルゴリズムを構成している。ただし、編集操作を core に限定したとしても、そのような XPath 式が効率良く得られるかは自明でない。そこで本論文では、第 5 章において xd-graph という新しいグラフを提案している。Xd-graph の重要な点は、入力 XPath 式を修正して得られる正しい XPath 式がグラフ上の開始ノードから受理ノードへの経路に対応し、かつ、開始ノードから受理ノードへ至る経路 p の重みが「入力 XPath 式を p が表す XPath 式に修正するためのコスト」に一致していることにある。したがって、入力 XPath 式に最も近い K 個の XPath 式を得るには、xd-graph 上で K 最短経路問題を解けばよいことになる。第 6 章では、この考えに基づいて、DTD の下で入力 XPath 式に最も近い K 個の XPath 式を求めるアルゴリズムを構成し、本アルゴリズムが simple な XPath 式に対して多項式時間で動作することを示している。また、XP に属する XPath 式に対して、述語のネストの深さが定数で抑えられている場合、本アルゴリズムが多項式時間で動作することも示している。さらに、適切な枝刈り処理を行うことにより、本アルゴリズムが実際の DTD の下で効率よく XPath 式を修正可能であることを、評価実験によって明らかにしている。

正規木文法は、DTD より真に表現力の高いスキーマ言語である。第 7 章では、正規木文法にも対応できるように上述のアルゴリズムを拡張している。DTD と正規木文法との最も大きな違いは、後者は一つの要素名に対して複数の型を割り当て可能であることである。まず、この性質を xd-graph に採り入れて拡張し、xd-graph と同様の性質を有する xg-graph を構成している。次に、この xg-graph を用いて、正規木文法の下で入力 XPath 式に最も近い K 個の XPath 式を求めるアルゴリズムを構成している。さらに、このアルゴリズムの計算複雑さについても考察し、DTD の場合と同様の効率で提案アルゴリズムが動作可能であることを明らかにしている。

## 審査の要旨 (2,000 字以上)

### 【批評】

これまで、関係データベースにおいては、問合せ言語 SQL の記述を支援するための手法が提案されてきた。一方、XML は関係データベースと比べて複雑なデータ構造を扱えることから、問合せ式の記述を支援するための手法を構成するのはより困難かつ挑戦的な課題である。本研究では、XML において最も普及している問合せ言語 XPath を対象に、スキーマを参照して XPath 式の修正を行う問題 (XPath 式修正問題) について考察している。

本問題は静的解析の一種であるが、XML における静的解析は計算困難な問題が多く、本研究が対象とする XPath 式修正問題も効率よく解けるか否かは自明ではない。本研究の特徴は、アルゴリズム・計算複雑性の観点から、XPath 式修正問題の計算困難な部分と効率よく解ける部分を明らかにしているところにある。論文の構成は、まず、第 2 章で関連研究を概観している。第 3 章から第 4 章にかけて、XPath 式の 2 つのクラスおよび XPath 式に対する編集操作の 2 つのクラスを定義し、これらクラスの観点から XPath 式修正問題の計算困難性を論じている。第 5 章から第 6 章にかけて、DTD の下で入力 XPath 式に対する修正候補を発見するアルゴリズムを提案し、その時間計算量を求めている。第 7 章は、第 6 章のアルゴリズムを正規木文法に適用できるように拡張している。第 8 章では、提案アルゴリズムの評価実験を行っている。第 9 章では、これまでの議論を踏まえて、本論文の到達点を明らかにしており、最後に第 10 章でまとめている。以下、各章毎に議論を概観し批評を行う。

第 1 章は、XML は RDB と比べて複雑な構造を有する一方、RDB と比較して問合せ式の記述を支援する手法が非常に少ない現状を指摘し、XPath 式修正問題を考察することの妥当性について述べている。さらに、XML における静的解析は多くの場合に計算困難であり、本論文の対象とする XPath 式修正問題が効率よく解けるかが自明でないことを述べている。本論文が目標としている XPath 式修正問題の計算複雑さを解明するという課題の妥当性は十分に説明されており、研究の意義は十分に認められる。

第 2 章は、上記研究課題の関連研究について述べている。まず、文字列や木構造など、種々のデータに対する編集操作について概観・整理している。次いで、XML における静的解析問題、および、XML における問合せ式の修正に関する研究を概観・整理し、本研究の位置付けを明らかにしている。本論文をまとめる範囲で十分な調査をしていると言えるが、XPath と述語論理やオートマトン等との関連が述べられていれば、本研究の位置付けを別の観点からも示し得たと考えられる。

第 3 章および第 4 章は、XPath 式の 2 つのクラス simple と XP を定義するとともに、XPath 式に対する編集操作の 2 つのクラス core と extended を定義している。そして、これら編集操作および XPath のクラスが、XPath 式修正問題の計算複雑さに影響を与える主要な要因であることを明らかにしている。まず、編集操作として extended を許した場合、XPath 式を simple なものに限定しても XPath 式修正問題が NP 困難であることを示している。さらに、extended の編集操作のうち 1 つのみを許した場合でも、XPath 式修正問題は依然として NP 困難であることを示している。これらの議論により、XPath 式修正問題の計算困難な部分を明確に示すことができたと言える。

第 5 章および第 6 章は、XPath 式の修正候補を列挙するアルゴリズムを提案し、その時間計算量を明らかにしている。提案アルゴリズムは、編集操作を core に限定した上で、DTD の下で入力 XPath 式に最も近い K 個の XPath 式を求めるものである。本アルゴリズムの特徴は、入力 XPath 式を修正して得られる正しい XPath 式の集合を、xd-graph というグラフで表現していることにある。これにより、効率よく入力 XPath 式に最も近い K 個の XPath 式を求めることが可能となっている。実際、本アルゴリズムが simple な XPath 式に対して多項式時間で動作することを評価実験で確認し、さらに XP に属する XPath 式に対しても、述語のネストの深さが定数で抑えられている場合には多項式時間で動作することも示している。このアルゴリズムにより、XPath 式修正問題において効率よく解ける部分を明らかにすることに成功しており、その意義は大きいと言える。

第 7 章は、上記のアルゴリズムを正規木文法の下で動作するように拡張している。正規木文法には、一つの要素名に対して複数の型を割り当て可能であるという DTD には無い特徴がある。この特徴を xd-graph に採り入れ、xd-graph と同様の性質を有する xg-graph を構成し、このグラフ上での探索を行うことにより入力 XPath 式に最も近い K 個の XPath 式を求めるアルゴリズムを実現している。さらに、本アルゴリズムの計算複雑さについて考察し、DTD の場合と同様の効率で本アルゴリズムが動作可能であることを明らかにしている。正規木文法は、W3C XML Schema や RELAX NG の形式モデルとして用いられている。この結果は、本アルゴリズムが XML の主要なスキーマ言語に適用可能であることを示す有意義な成果であると言える。

第 8 章は、アルゴリズムに関する評価実験について述べている。適切な枝刈り処理を行うことにより、本アルゴリズムが実 DTD の下で効率よく XPath 式を修正可能であることを明らかにしている。実験に用いた DTD(auction.dtd)は実 DTD の中ではサイズも大きく複雑なものであるため、本アルゴリズムは他のスキーマに対しても効率よく動作すると期待される。ただし、より多くのスキーマを用いて評価実験を行えば、本アルゴリズムの評価をより確かなものにできたと思われる。

第 9 章の考察、および第 10 章の結論では、第 1 章で示した本論文の課題に対して、本研究の取り組みが明らかにした範囲を論じている。本研究の遂行によって設定した課題が解決できていることを示すとともに、新たに明らかとなった課題の存在も示唆している。

以上を総合的に判断すると、本論文は情報学の学位論文として十分な内容を含んでいると認められる。

### 【最終試験結果】

平成 28 年 7 月 1 日、図書館情報メディア研究科学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程（課程博士）の学位論文審査に関する内規」第 23 項第 3 号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

### 【結論】

よって、本学位論文の著者は博士(情報学)の学位を受けるに十分な資格を有するものと認められる。