論文内容の要旨

博士論文題目

XML 文書の部分文書検索に関する研究

氏　名

絹谷 弘子

(論文の要旨)

構造化文書の一種である XML(Extensible Markup Language)は，ネットワーク上のデータや文書の交換手段として利用されるだけではなく，電子新聞記事や電子政府の各種文書などの公開文書に利用されるなど，その利用範囲は急速に拡大している．それに伴い，必要な情報を効率良く高速に検索できる XML 文書を対象とした検索エンジンの必要性も高まっている．現在の検索エンジンは，いくつかのキーワードを問合せとして入力し，関連性の高い文書に関する情報を検索結果として表示する．しかし，問合せ結果は，ネットワーク上の流通単位であるファイルであり，たとえ問合せとの関連性が高い部分がファイル中の一部分であったとしても，その部分を対象とした検索や結果の表示を行えない．

本研究の目的は，利用者に負担のない問合せの入力による適切な部分文書の検索と表示である．本研究で求める部分文書は，検索対象文書の論理木構造を保持した形の部分文書とする．各部分木の根ノードと部分文書が一対一に対応する．部分文書検索においては，部分文書の文書内容と同様にこの部分文書までの元文書中における論理構造上の位置を示す経路式が表す文脈情報を得ることが重要であるという考えに基づく．XML 問合せ言語では，この経路式を利用者が指定することを前提としているが，一般の利用者は，文書中の特定の場所を示す方法を持たない．利用者が検索結果の絞り込みや検索結果の再利用をするためにも部分文書の位置を表す経路式と部分文書の内容を利用者に提示することが重要である．さらに，従来の KWIC による表示では，検索結果として入力キーワードの周辺の文を表示することによって利用者はキーワードが使われている文脈を知ることができたが，XML 部分文書検索においても，検索結果として入力キーワードの周辺の部分文書と，その部分の構造上の位置を経路式として表示することにより利用者に対してより多くの文脈情報を提供できる．

本論文では，まず，利用者が検索対象となる XML 文書集合の文書構造に対して持っている知識の程度を次の二つに分けている．一つ目は，利用者が標準語彙の構造知識を持つ場合であり，二つ目は，文書構造に関する知識を全く持たない場合である．その上で，それぞれの場合に対応した問合せを，構造とキーワードの組，および，キーワードのみとしてモデル化し，問合せ条件を満たす最適な部分文書を求めるアルゴリズムを提案し，その有効性を実験によって確認した．本研究によって，利用者が文書構造についての知識が乏しく，問合せ言語の複雑な構文を知らない場合でも，必要な文書の一部とその文脈情報を取り出し，表示することができた．

氏　名	絹谷弘子

（論文審査結果の要旨）

平成13年12月25日に開催した公聴会の結果を参考に，平成14年2月18日に本博士論文の審査を行った．

絹谷弘子は，本博士論文においてXML文書の部分文書検索を取り上げ，部分文書検索モデルを定義し，検索の前提条件を，問合せを行う利用者の文書構造に対する知識によって二種類に分類してこの問題を論じている．

まず，一番目にXML部分文書検索モデルを定義するにあたり，XML文書検索に対する要求を分析し，部分文書検索をXML検索エンジンの要素技術として位置付け，構造とキーワード指定によるXML部分文書検索を論じている．XML文書の文書構造から文脈の切れ目を抽出することによってKWIC(KeyWord In Context)を可能とする部分文書検索を提案している．ここで提案されている部分文書検索は，XML問合せ言語や，従来の情報検索では扱えない検索要求に答えることが可能であることを，単純問合せ，ブーリアンAND問合せとOR問合せにおいてそれぞれの問合せの意味とその検索アルゴリズムを論じ，評価実験を行うことで示す．

二番目は，キーワード指定によるXML部分文書検索を，すべての部分文書を対象とした場合と構造上の文脈の切れ目をヒューリスティックに求めた文脈ノードに対応した部分文書を対象とした場合について比較し，本論文で提案された後者の手法によって算出される適合率，再現率が向上することを実験により確認している．その結果，従来の研究においてはできなかった，最適な粒度の部分文書を自動的に検索結果とすることが文書構造と文書内容を考慮した類似度を利用することで可能となることを示している．

これらの研究は，検索結果から文書構造と文書内容の文脈情報を得ることを目的としたXML部分文書検索の研究であり，今後の研究の基盤を形成する重要な研究成果である．

以上により，本博士論文は，今後重要性を増すと思われる構造化文書の部分検索のための基盤技術の構築に貢献しており，学術上，実用上寄与するところが少なくない．よって本論文は，博士（工学）の学位論文として価値があるものと認める．