

平成 17 年度 情報工学コース卒業研究報告要旨

吉川 研究室	氏 名	林 由 紀 子
卒業研究題目	構造化文書を対象とした 自然言語検索に関する研究	
<p>近年，XML が広く普及してきたことにもない，構造化文書の検索がますます重要になっている．構造化文書を対象とした情報検索は，従来のプレーンテキストを対象とした情報検索と比較して，次のような利点がある．</p> <ol style="list-style-type: none">1. キーワードだけでなくそれが含まれる要素名も指定して検索ができる2. 文書単位の検索に加えて，部分文書単位の検索もできる <p>一般に，文書の構造を活用した検索を行うためには，XPath などの XML 用の問い合わせ言語を用いる．しかし，問い合わせ言語を用いた検索は，問い合わせ言語の構文規則についての知識，および，XML のタグセットのような検索対象の文書構造についての知識を必要とし，一般の利用者が用いるには適さない．</p> <p>本論文では，自然言語で記述されたクエリを問い合わせ言語に変換する手法を提案する．これにより，利用者は自然言語文を入力しさえすればクエリを作成でき，問い合わせ言語の構文規則についての知識，検索対象の文書構造についての知識は不要となる．</p> <p>本手法では，入力された自然言語のクエリを解析し，パターンマッチングを用いて問い合わせ言語の生成に必要な情報を抽出する．抽出した情報は，要素同士，あるいは，要素とキーワードの関係を表す述語表現に変換する．例えば，“[要素 X] [containing] [一般名詞 A]” という語の並びパターンは，要素 X はキーワード A を含むという意味の述語表現 $\text{about}(X,A)$ に変換する．このようなパターンマッチングにより，複数の述語表現を生成する．これらの述語表現を個別に問い合わせ言語に変換し，それらを 1 つに統合して最終的な出力とする．</p> <p>本手法の有効性を評価するために，自然言語で記述されたクエリを問い合わせ言語に変換する実験を行った．実験には，INEX2005 のクエリセット 47 件を用いた．これらのクエリセットでは，情報要求が自然言語と問い合わせ言語の両方で記述されている．評価は，システムが生成した問い合わせ言語と，クエリセットに記述されている問い合わせ言語を比較することにより行った．実験の結果，本手法の有効性を確認した．</p>		