구조 변환과 엘리먼트 매칭에 의한 XML 스트림 질의 처리 설계

이 호 석
호서대학교 공과대학 뉴미디어공학과
hslee@office.hoseo.ac.kr

A design of XML stream querying using structure transform and element matching

Ho Suk Lee
New Media Department, College of Engineering, Hoseo University

요 약

XML 스트림에 대한 질의 처리 방식으로는 유한자동기계(finite state automata) 방식이 많이 연구되었다. 대표적인 방식으로는 YFilter가 있다. 그러나 YFilter 방식은 경로 노드(path node), 질의어 인덱스(query index), 그리고 유한자동기계를 구성하고 계속 유지해야만 하는 문제가 있다. 본 논문에서는 XML 스트림에 대한 질의 처리 방식으로 문서 구조 변환과 엘리먼트 패턴 매칭 방식을 제안한다. XML 문서를 SAX(Simple API for XML) 메소드를 사용하여 quadruple 배열로 변환하고 질의어 경로표현식(path expression)도 엘리먼트 문자열로 변환한다. 다음에 경로표현식의 엘리먼트와 quadruple 배열의 엘리먼트를 패턴 매칭하여 XML 스트림에 대한 질의어 처리를 구현하였다. 간단한 실험 결과를 제시하여 제안된 방법의 가능성을 보여주었다.

. 서론

.1 연구 동기

최근 스트림 데이터에 대한 활발히 연구가 진행되고 있다. 스트림 데이터란 데이터 요소가 연속적으로 발생하는 데이터로 판매 내역 데이터, 신용카드 결제 데이터, 시스템 및 네트워크 로그 데이터, 센서 네트워크 데이터, 웹 처리 데이터 등이 스트림 데이터의 예이다. XML은 웹에서 시스템과 사용자의 데이터 표현과 교환 방식으로 표준화 되었다. 그리고 최근에 XML 방식으로 표현된 스트림 데이터도 많이 발생하고 있다. 이들 스트림 데이터는 전통적인 방식으로 데이터베이스에 저장하고 질의 처리를 하는 것이 적절하지 않다. 왜냐하면 일반적인 데이터베이스는 연속적으로 발생하는 많은 양의 데이터를 저장하고 처리하는 방식으로 설계되지 않았으며 연속 질의 처리를 직접적으로 지원하지도 않기 때문이다. 또한 이러한 스트림 데이터에는 근사적이고 융통성 있는 질의 처리가 필요한 경우도 있다. 스트림 데이터를 처리하는 시스템을 DSMS(Data Stream Management System)라고 한다[1]. 본 논문에서는 XML 문서 스트림에 대하여 엘리먼트 패턴 매칭을 이용한 질의어 처리 방식을 제안한다.

1.2 관련 연구

XML 스트림에 대한 질의어 처리는 이벤트에 의한 유한자동기계(FSM, Finite State Automata) 방식이 많이 연구되었다[2]~[7]. 이 방식에서는 질의어의 경로표현식(Path Expression)을 유한자동기계로 매핑하고 XML 스트림을 SAX 파서의 메소드로 처리하면서 발생하는 엘리먼트에 의하여 유한자동기계의 상태가 옮겨져서 최종적으로 accept 상태에 도달하면 해당 질의어에 대한 처리가 완료된 것으로 간주하였다. 이 방식으로서는 각각의 질의어를 DFA(Deterministic Finite Automata)로 구현한 XFilter가 있고[3], 이 방식을 개선하여 여러 개의 질의어에 대한 DFA를 병합하여 NFA(Non-deterministic Finite Automata)를 구현한 YFilter가 있다[4][5][6]. 또한 참고문헌 [7]에서는 XSM(XML Stream Machine) 방식의 XML 스트림 질의어를 구현하여 XML 스트림에 대한 질의 처리를 설명하였다. 참고문헌 [7] 에서도 DFA 방식에 의한 XML 스트림에 대한 질의어 처리를 보여주었다. 참고문헌 [8]에서는 타이프 시스템을 정의하여 XML 스트림에 대한 질의어 처리를 해결하였다. 이 논문에서는 XML 스트림에 대한 질의어 처리를 2 단계로 처리하였다. 참고문헌 [9]에서는 역 인덱스를 사용하여 XML 스트림을 필터링한 방식을 설명하였다. 참고문