

인용문헌의 효율적인 검색을 위한 XML기반 저장모델 설계 (Design of SCI Repository Model based on XML for the Efficient Retrieval)

이계준, 조현양, 최재황, 손강렬, 윤희준
한국과학기술정보연구원

{kjlee, hycho, findit, krshon, k2@kisti.re.kr}

Kye-Jun Lee⁰ Hyun-Yang Cho, Jae-Hwang Choi, Kang-Ryul Shon, Hee-Jun

Yoon

Korea Institute of Science and Technology Information

요 약

급속히 발전하고 있는 웹 상에서 또는 컴퓨터 상에서 다양하고 많은 양의 원문들에 대한 모든 정보를 효율적이고 체계적으로 축적, 검색, 관리하여 이용자들에게 능동적인 서비스를 제공할 수 있는 시스템의 필요성이 대두되었다. 또한, 원문정보를 효과적으로 관리하고 서비스하기 위하여 플랫폼 독립적이며, 문서정보의 전송과 교환이 용이하고, 문서가 나타내는 의미를 그대로 표현할 수 있는 기술적인 해결책이 필요하게 되었다.

본 논문에서는 원문정보를 검색할 때에 원하는 결과를 얻기 위하여 많은 조인(join)연산, 필터링(filtering), 재구성을 거쳐서 결과를 획득할 수 있으며, 문서의 내용정보에만 중점을 두어 문서를 저장 관리하던 기존의 시스템의 문제점들을 보완하기 위하여 XML을 기반으로 하는 저장모델을 제안하고자 한다.

본 논문의 저장 모델은 범용적인 측면의 마크업(Markup)인 XML을 이용하여 원문을 XML문서로 변환하여 작성한 다음 파싱 과정을 통해 원문의 구성요소와 속성 및 논리적인 구조를 분석하고 효율적인 관리와 검색을 위하여 DOM Tree를 기반으로 정보를 재구성하여 모든 데이터 및 논리적인 정보를 손실 없이 저장하기 위한 것이다. 또한 같은 정보의 구성요소들은 같은 레벨(level)에 형제(sibling) 관계를 가지도록 위치시켜 데이터베이스 구축을 용이하게 하며, Path Encoding 방식의 색인 기법을 통하여 유일(Unique)한 코드 생성으로 코드만으로도 상하 계층관계를 유추할 수 있도록 설계하여, 검색 대상에 따른 검색 위치 선정을 가능하도록 하였으며, 구성요소들 간의 많은 참조나 링크가 있을 때에 발생할 수 있는 순환적인 참조 관계를 해결 할 수 있는 저장모델 설계를 목적으로 한다.