
HTML 문서의 무결성 유지 시스템의 설계 및 구현

조이기 · 이영운 · 황인문 · 양수영 · 김원중

순천대학교

The Design and Implementation of HTML Document Integrity Management System

Lee-gi Cho · Young-Woon Lee · In-Moon Hwang · Su-Young Yang · Won-Jung Kim

Sunchon National University

E-mail : dbcho,ywe,him,ysy,kwj@sunchon.ac.kr

요 약

KLDP(리눅스 한글 문서 프로젝트, <http://kldp.org>)와 같이 많은 분량의 HTML 문서들로 이루어진 매뉴얼 사이트를 관리하는데 가장 어려운 점은 현수 참조(dangling reference), 오문 참조(inaccurate content reference)와 같은 깨진 링크(broken link)를 관리하는 것이다.

본 논문에서는 웹 사이트의 HTML 문서들 사이에 존재하는 관계성(Relationship)과 제약 조건(Constraints)을 정의하여, HTML 문서의 삽입, 삭제, 변경이 발생할 경우 사용자에게 통지하거나, 트리거를 수행하여 HTML 문서들의 무결성(Integrity)을 유지할 수 있도록 하는 HIMS(HTML Document Integrity Management System)시스템을 설계 및 구현하였다.

ABSTRACT

It is difficult to manage broken link with dangling reference, inaccurate reference in the manual site that is consist of HTML documents of much quantity as KLDP(Korean Linux Documentation Project, <http://kldp.org>) Web site.

In this paper, we define relationship and constrain conditions that exist between Web site's HTML documents. And we design and implement HIMS(HTML Document Integrity Management System), which notify user that integrity violation happens or launch trigger operation to keep integrity between HTML documents in case of insert, delete, update.

키워드

HTML, Constraint Condition, Trigger, Integrity

1. 서 론

KLDP(Korean Linux Documentation Project, <http://kldp.org>)와 같은 많은 매뉴얼 사이트나 연구 결과 정보 제공 사이트들은 일반적으로 많은 HTML 문서들로 이루어져 있으며 각 문서들간의 연결은 HTML의 Anchor 태그를 사용한다. 개인의 홈페이지의 경우는 문서의 양이 많지 않고 개인과 연관된 문서이기 때문에 관리가 잘 되어 있다. 그러나 KLDP와 같은 다수의 사람이 관리·운영하는 사이트의 경우에는 현수 참조(dangling reference)와 오문 참조(inaccurate content reference)와 같은 깨진 링크(broken link)를 관리하는데 많은 어려움이 따른다. 일반적으로 HTML 문서의 내용이 변경되거나 삭제가 된 경우 관리자는 정해진 규칙에 따

라 조치를 취한다. 그러나 HTML 문서의 저자나 그 외의 사람이 임의대로 HTML 페이지의 내용 중 연결 구조에 관련된 부분을 변경하거나 삭제하는 경우 문서들간의 연결 구조에 대한 무결성(Integrity)을 보장하기가 어렵다. 데이터베이스 시스템에서는 데이터의 무결성을 유지하기 위하여 개체(Entity) 사이의 관계성과 제약 조건을 정의하여 트리거를 사용하여 임의의 사용자가 데이터베이스의 상태를 잘못된 의미를 가진 데이터로 변경하는 경우를 제약하고 있다.

본 논문에서는 데이터베이스 시스템에서 사용하는 관계성과 제약 조건의 정의를 HTML 문서들 사이에도 적용하여 HTML 문서들 간의 연결

구조의 무결성을 유지할 수 있는 HIMS(HTML Document Integrity Management System)을 설계 및 구현하였다.

II. 관련 연구

2.1 데이터베이스에서의 무결성

데이터베이스에서의 무결성은 데이터의 정밀성(accuracy)이나 정확성(correctness)을 의미하는 것으로서 데이터베이스에 저장된 데이터를 의미적으로 정확하게 유지하는 것을 말한다. 무결성 제약 조건(Integrity Constraints)은 데이터베이스에서 어떤 데이터 값이 허용되고 어떤 트랜잭션의 수행 결과가 인정되는지를 나타낸다. 예를 들어 나이는 1부터 150사이의 정보값이어야 한다는 데이터 값에 대한 무결성 제약 조건이 정의될 수 있고, 나이를 감소시키는 트랜잭션의 결과는 데이터베이스에 반영될 수 없다는 것 등이다. 또한 외래키(Foreign Key) 제약 조건은 외래키에 의해 참조되는 튜플(tuple)은 관계를 맺고 있는 릴레이션에 반드시 존재하여야 함을 명시한다. 따라서 무결성 제약 조건을 만족시키지 못하는 갱신(update)은 거절(Reject)되거나 트리거 연산을 호출하여 만족시키도록 하여야 한다[1,4].

2.2 데이터베이스 무결성 유지 서버 시스템

데이터베이스 시스템은 무결성을 유지하기 위하여 무결성 유지 서버 시스템을 운영한다. 무결성 유지 서버 시스템은 무결성 규정을 정의해 놓고 데이터베이스의 상태가 이 규정에 위반되지 않도록 관리하는 것으로서 무결성 규정에는 규정 이름, 검사 시기, 제약 조건, 위반 조치 등이 포함된다. 표 1.은 무결성 유지 서버 시스템이 관리하는 무결성 규정의 예로서 AFTER 절은 검사 시점이 STUDENT.YEAR의 갱신 직후라는 것을 명시하며, CHECK 절은 STUDENT.YEAR>0이 갱신 연산뒤에도 만족해야 할 제약조건이며, ELSE 절은 검사에 실패했을 때 규정 R1이 위반되었다는 것을 프린트하고 갱신 연산을 거부한다는 것을 명세하고 있다[1,2].

표 1. 무결성 규정

<pre> R1: AFTER UPDATING STUDENT.YEAR: CHECK(STUDENT.YEAR>0) ELSE DO; PRINT "R1 violated"; REJECT; END; </pre>

2.3 트리거(Trigger)

기본적으로 무결성 유지 서버 시스템에서는 무결성 규정이 위반되는 경우 해당 트랜잭션을 취소

한다. 그러나 해당 트랜잭션을 취소하는 것 이외의 조치가 필요한 경우에는 트리거를 사용한다. 트리거는 메시지를 보내거나, 어떤 값을 갱신하게 되면 자동적으로 다른 값을 갱신하기 위하여 무결성을 만족시키기 위한 프로시저를 작동시키는 경우에 사용하는 것으로서 어떤 조건이 만족되는 경우 실행해야 할 조치들을 명세 해 놓은 것이다[1,4].

III. HIMS

HIMS(HTML Document Integrity Management System)는 웹 서버의 HTML 문서들간의 연결 구조의 무결성을 유지하기 위한 시스템이다. HTML 문서들간의 연결 구조에 제약 조건을 설정하고 해당 제약 조건에 위배되는 경우 지정된 트리거를 발생시켜 HTML 문서들간의 연결 구조의 무결성을 유지하도록 하는 것이다. HIMS 시스템은 HIMS Server와 HIMS Client로 구성되며, HIMS Server는 웹 서버가 존재하는 시스템에 위치하며, GCC 2.8.1과 JDK 1.3.1로 개발되었다. 솔라리스 7 환경에서 작동되며, 호환되는 웹 서버는 Apache 1.3.22 이다. HIMS Client는 MS Visual Basic 6.0 으로 개발했으며 HIMS Server와는 소켓으로 데이터를 교환한다. 그림 1.은 HIMS 시스템 구조이다.

3.1 HTML문서들 사이의 관계성 및 제약 조건

HTML 문서들 사이의 관계성은 상위 페이지의 일부뿐만 바뀌어 새로운 하위 페이지를 생성하는 일반화(Generalization) 관계, 각각 독립적인 여러 개의 페이지들이 모여서 새로운 페이지를 생성하는 집산화(Aggregation) 관계, 그리고 결제 시스템의 흐름에서 볼 수 있는 페이지들 간의 연관(Association) 관계로 분류할 수 있으며, 이들에 적용되는 무결성 제약 조건은 다음과 같이 크게 4가지로 나눌 수 있다.

(1) When Parent Delete, Restricted

이 제약 조건은 부모 노드를 삭제할 때 자식 노드가 존재하는 경우 부모 노드를 삭제할 수 없도록 하는 제약 조건이다.

(2) When Parent Delete, Child Delete

이 제약 조건은 부모 노드가 삭제되는 경우 부모 노드와 연결되어 있는 자식 노드도 같이 삭제하는 경우이다.

(3) When Child Delete, Restricted

이 제약 조건은 자식 노드가 삭제되는 경우 부모 노드와 연결되어 있는 경우 자식 노드를 삭제할 수 없는 경우이다.

(4) When Child Delete, Parent Delete

이 제약 조건은 자식 노드가 삭제되는 경우 연결되어 있는 부모 노드까지 같이 삭제하는 경우이다.

3.2 HIMS Server

HIMS Server는 Site Parser, Site Monitor, HIMS Manager, HTML Repository, Trigger Process Repository, Site Information Database, Changed Site Information Database, Weak Relation Information Database, User Notifier, HIMS Client로 구성된다.

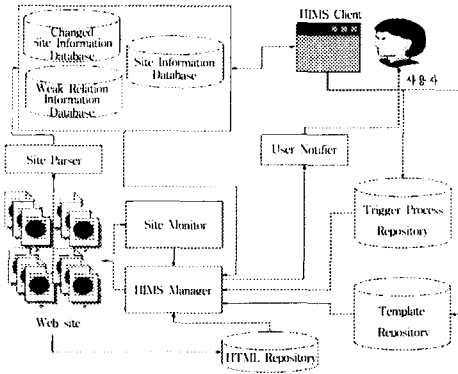


그림 2. HIMS 시스템 구조

User Notifier로 구성된다. Site Monitor는 GCC 2.8.1로 개발되었으며 나머지 모듈은 JDK 1.3으로 개발하였다.

(1) Site Parser

Site Parser는 웹 문서가 저장되어 있는 디렉토리에서 작동한다. 루트 노드(최상위 페이지)에 해당하는 HTML 페이지를 사용자가 지정하면, 지정된 HTML 문서를 기준으로 너비 우선 탐색 알고리즘을 이용하여 연결된 모든 HTML문서들을 파서한다.

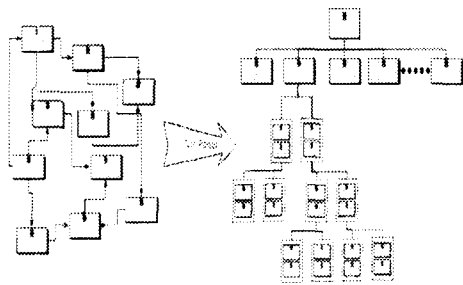


그림 3. Site Parser 변환 과정

HTML 문서에서 Anchor 엘리먼트를 파서하여 추출한 연결 정보는 자바에서 제공하는 벡터 자료형으로 구성된 Site Information Database에 저장한다. 저장되는 정보는 루트 노드와의 거리, 상위 노드의 주소, 현재 노드의 정보를 저장한다. 중복

된 노드인 경우 루트 노드와의 연결 거리가 가장 짧은 노드와의 연결을 선택하고 나머지 연결들은 Weak Relation Information Database에 저장한 후 제거한다. Site Parser의 최종 결과물은 그림 2와 같은 트리 구조로 구성되며, Site Parser의 작동 주기는 사용자가 결정한다.

(2) Site Monitor

Site Monitor는 일종의 데몬(Daemon)으로서 파일 시스템에서 웹 서버의 도큐먼트 디렉토리 하위 디렉토리 내에서 파일 생성, 삭제, 수정 및 디렉토리 생성, 수정, 삭제가 일어나는 경우 해당 파일 이름과 관련된 정보를 HIMS Manager에게 통보하고 Changed Site Information Database에 저장한다.

(3) HIMS Manager

HIMS Manager는 크게 두 가지 기능을 수행한다. 첫째, 파일 시스템에 변동이 생기는 경우, 해당 HTML 문서에 대한 제약 조건을 추출하여 해당하는 Trigger를 작동한다. 제약 조건이 설정되어 있는 HTML 문서가 수정, 변경인 경우에는 Trigger Process Repository에 저장되어 있는 지정된 Trigger를 작동한다. 제약 조건이 설정되어 있지 않은 페이지가 삭제된 경우는 Template Repository에 저장되어 있는 지정된 HTML문서로 대체한다. 삭제 불가인 경우에는 삭제된 HTML 문서나 디렉토리를 HTML Repository에서 찾아 복원시키고 변경 정보는 Changed Site Information Database에 저장한다. 둘째, HIMS Client와의 소켓 연결을 통하여 사용자에게 변경된 정보를 제공하고, 사용자가 지정한 제약조건이나 사용자가 만든 Template 문서를 전송 받는다.

(4) HTML Repository

HTML Repository는 웹 서버의 도큐먼트 디렉토리를 그대로 복사하여 가지고 있는 저장소이다. 파일이 삭제되는 경우 HIMS에서 처리할 수 없기 때문에 삭제된 후 다시 복원시키기 위하여 웹 서버의 웹 문서가 저장되어 있는 폴더와 동일한 HTML 문서들을 저장한다.

(5) Trigger Process Repository

Trigger Process Repository는 트리거들의 저장소이다. HIMS는 제약 조건이 설정되어 있는 HTML 문서에 변동이 생기는 경우 해당 트리거는 Trigger Process Repository에서 호출되어 실행된다.

(6) Template Repository

트리거에 의해 대체되는 페이지들이 저장되어

있는 Repository로서 해당 페이지가 삭제되거나, 변경이 되는 경우 사용된다. 기본적으로 제공하는 페이지 외에 사용자가 직접 작성한 문서도 저장된다.

(7) Site Information Database

Site Parser에 추출된 웹 사이트 구조를 저장하고 있는 벡터 기반의 데이터베이스이며, 웹 사이트의 계층적 구조 외에도 각 페이지 별로 설정되어 있는 연결 관계와 제약 조건이 저장되어 있다.

(8) Changed Site Information Database

입력의 사용자에게 제약 조건이 설정되어 있는 HTML 문서들이 변경되는 경우 HIMS는 트리거를 작동하여 병세된 행위를 수행한다. 이런 경우 Site Information Database에는 트리거가 수행된 이후의 정보만을 저장하고 있으며, Changed Site Information Database에는 이전의 정보들을 보존하고 있다.

(9) Weak Relation Information Database

연결 관계는 주 관계(Strong Relation)와 보조 관계(Weak Relation)가 있으며, 주 관계는 루트 노드와 부모 노드와의 거리가 가장 짧은 노드와 설정되며, 나머지 연결 관계는 보조 관계로 설정된다. 보조 관계로 연결되어 있는 HTML 문서들 간의 연결 정보와 제약 조건은 Weak Relation Information Database에 저장한다. 주 관계에 연관된 제약조건과 트리거가 수행되고 난 후 보조 관계의 제약 조건과 트리거가 수행된다. 주 관계와 보조관계의 제약 조건이나 트리거의 등급은 동일하다.

(10) User Notifier

Trigger에 저장되어 있는 정보를 수행한 경우 사용자에게 E-Mail로 저장되어 있는 정보를 알려주는 시스템이다.

3.3 HIMS Client

HIMS Client는 MS Visual Basic 6.0으로 구현되었으며 HIMS Server와는 소켓으로 통신한다. HIMS Manager에게 전송받은 Site Information Database, Changed Site Information Database, Weak Relation Information Database 등의 정보를 사용자에게 전달하고, 제약 조건의 설정, 사용자가 지정한 트리거 및 Template 페이지를 HIMS Manager에게 전송한다. 그림 3은 HIMS Client의 화면이다.

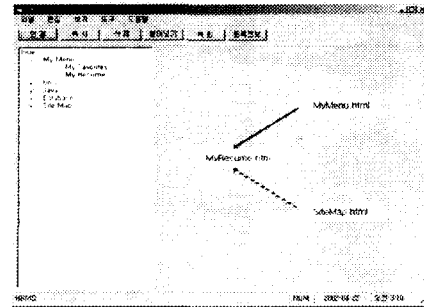


그림 4. HIMS Client 화면

IV. 결 론

본 논문에서 제안한 HIMS 시스템은 HTML 문서들 간에 제약 조건을 설정하여 해당 제약 조건에 위배되는 이벤트가 발생하는 경우 해당 트리거를 작동시켜 HTML 문서들 간의 연결 구조의 무결성을 유지하도록 하였다.

앞으로의 연구 과제는 HTML문서가 이동되거나, HTML 문서의 Anchor의 엘리먼트 정보가 변경되는 경우 처리할 수 있도록 해야 하며, HTML 문서들 간의 연결 구조뿐만 아니라 웹 어플리케이션에서 적용할 수 있도록 하는 것이다.

참고문헌

- [1] 이석호, 데이터베이스시스템, 정익사, p.427-433, 1997
- [2] 유남현, 김원중, 웹 사이트 목차 디렉토리 생성 에이전트, 한국정보과학회, 권 호.p. -. 2000
- [3] Nation, D., Plaisant, C., Marchionini, G., Komlodi, A., "Visualizing Websites using a Hierarchical Table of Contents Browser: WebTOC", 1997
- [4] Stefano Ceri, G.Pelagatti, "Distributed Database Principles&Systems," McGraw-Hill, P.6 1-62, 1984