가상 문서의 효율적인 관리를 지원하는 대용량 링크정보 관리서버의 개발

김영승[0] 박인석 현순주
한국정보통신대학원대학교 공학부
(seung92[0], ispark, shyun)@icu.ac.kr

Design and Implementation of Link Server for Efficient Management of Virtual Documents

Young Seung Kim[0], In Suk Park, Soon J. Hyun
School of Engineering.
Information and Communications University (ICU)

요 약

지식기반 사회에서 보다 효율적인 정보이용을 지원하기 위해 디지털 도서관 시스템인 MIRAGE가 구현되어졌다. 인터넷이 비약적으로 성장하고, 접속하는 사용자의 수가 증가함에 따라 MIRAGE 시스템에서 관리하는 링크정보는 폭발적으로 증가하게 되었다. 따라서 링크정보를 검색할 경우 웹 정보 객체간의 다양한 링크정보를 이용하여 사용자 중심의 서비스를 제공하도록 요구되어지고 있다. 여기서 제안하는 링크서버는 링크가 가지는 유용한 정보를 체계적으로 분석하고 관리하여 검색에 사용함으로써 효율적인 검색 서비스가 가능하도록 제공하고 있다. 사용자는 자유롭게 웹상의 정보에 대해 링크정보를 만들고, 만들어진 링크정보는 데이터베이스에서 관리가 이루어진다. 하지만 웹상의 링크정보는 삭제가 되기나 변경이 될 경우가 있다. 여기에서는 시스템의 신뢰성을 향상시키기 위해서 링크서버는 따로 링크정보의 유효성을 주기적으로 검사를 하는 에이전트를 두고있다. 에이전트는 데이터베이스에 저장 되어진 링크정보가 항상 유효하도록 주기적인 검사를 통하여 항상 신뢰할 수 있는 링크서버가 동작이 되도록 지원을 하고 있다.

1. 서 론

가상문서를 기반한 디지털 라이브러리를 위해 MIRAGE 시스템에서는 링크정보를 가지고 운영이 되어졌다[4]. 링크정보를 위해 개발된 링크서버는 가상문서에 사용하는 링크정보를 데이터베이스에 저장하도록 설계가 이루어졌다. 링크정보는 내포링크, 참조링크로 나누어서 저장이 이루어졌다. 현재 MIRAGE 시스템은 블록 단위로 내포링크와 참조링크를 관리함으로써 데이터베이스에서는 블록에 관련된 정보를 함께 가지고 있다. 링크서버는 인터페이스를 통하여 추가, 삭제, 변경이 가능하도록 제공하고 있다.

가상문서를 기반하여 만드는 디지털 라이브러리에서 링크를 사용하는 이유는 여러가지의 이점이 있기 때문이다. 링크는 대용량의 멀티미디어가 차지할려는 공간을 절약할 수 있다. 여러 버전으로 표현이 가능하며 같은 내용을 서로 다른 미디어 형식으로 문서 표현이 가능하다. 가상문서는 실제로 문서를 전송하는 것이 아니라 링크 정보만을 전송하여 네트워크 트래픽을 감소시키므로 성능 향상을 추구할 수 있다. 이제 단순히 만들어진 정보에 의존하는 것이 아니라, 기존의 정보에 대해 링크를 통하여 개인 맞춤의 문서 제작이 가능하다. 따라서 링크는 여러가지 목적으로 활용될 수 있다.

MIRAGE 시스템에는 여러 콤포넌트가 상호 유기적으로 동작이 이루어지고 있고, 그 중의 하나가 바로 링크서버이다. 링크서버는 가상문서에서 사용하는 링크와 관련된 정보를 독립적으로 저장하고 관리하는 일을 한다. 즉, 새로운 가상문서가 저장되면 그와 관련된 링크정보를 데이터베이스에 저장하고, 필요한 링크정보를 얻기 위해 검색서버나 사용자 에이전트에서 정보 요청이 있을 경우에 이를 구조화하여 개별적으로 전달하는 역할을 하는 것이다. 이제 링크정보는 단순히 파일에 저장하는 것이 아니고 데이터베이스에 저장을 통하여 링크정보를 보다 효율적으로 저장을 하고, 신속하게 원하는 정보가 추출되도록 요구 되어지고 있다. 추가적으로 데드링크 검사 에이전트를 통하여 주기적으로 링크에 대해 유효성을 검증을 하여 항상 신뢰성을 가지는 MIRAGE 시스템이 되도록 한다.

본 보고서의 구성은 다음과 같이 이루어져 있다. 2장은 링크서버에 관련된 연구에 대해 소개를 한다. 3장은 링크서버를 위한 설계부분을 설명한다. 4장은 링크서버에서

사용하는 링크 관리 응용 프로그램에 대해 설명을 한다. 마지막 5장은 본 논문에 대한 결론을 내린다.

2. 관련연구

정보화 사회로 발전함에 따라 정보를 전자 문서화하여 효율적으로 관리하기 위한 요구가 급증하고 있다. 최근에는 흩어져 있는 문서에서 필요한 정보를 추출하여 가상문서로 이루어진 디지털 라이브러리가 사용되고 있다[1][2][3]. 저작도구를 통하여 기존에 존재하는 문서에서 필요한 정보를 사용하기 위해 링크정보를 이용한 가상문서의 제작으로 재사용성이 증대된다. 가상문서는 주로 XML 형식을 따르고 있다. 이제는 만들어진 가상문서에 포함되는 링크정보에 대한 효율적인 관리방법에 대해 연구를 진행하고 있다.

XML은 W3C가 인터넷 상에서 구조화된 전자 문서를 처리하기 위한 표준으로 발표를 한 문서 형식이다. 최근에 연구되어지는 가상문서는 XML을 기반으로 작성이 되어지고 있다. XML로 만들어진 가상문서는 링크정보와 같은 의미있는 정보에 대해 효과적으로 표현이 가능하다. 또한 XML을 저장하는 방법론에 대해서도 많은 연구가 이루어지고 있다. XML 문서의 구조정보를 데이터베이스에 저장하는 방식으로 크게 분할 저장 기법과 가상 분할 기법으로 나눌 수 있다. XML 문서를 저장하기 위해서는 기존 데이터베이스 시스템을 이용하는 방식과 별도의 데이터베이스 시스템을 개발하는 방식으로 나눌 수 있다.

3. 링크 서버 설계

3.1. 개념

그림 3.1은 링크서버와 MIRAGE 시스템과의 관계를 표현하고, 링크서버가 가리키는 Web 문서에 대한 링크정보가 유효하도록 관리가 이루어지고 있는 것을 나타내고 있다. MIRAGE 시스템에서 링크정보를 저장하도록 링크서버에 전달하고, 링크서버는 필요한 링크정보를 추출하여 MIRAGE 시스템으로 전달한다. 추가적으로 링크정보가 유효하지 않을 경우, 오류사항을 MIRAGE 시스템에 알려준다. 링크정보는 웹상의 문서에 대해서 일부 영역을 가리키므로, 간혹 링크정보가 유효하지 않을 경우가 발생하게 된다. 링크서버는 주기적으로 링크정보의 유효성을 검사하고 변경시 변경된 정보를 데이터베이스에 저장을 하게 된다. 링크정보의 유효성을 유지하기 위해 데드링크 검사 에이전트가 전담하고 있다.

그림 3.1 링크서버의 전반적인 기능

3.2. 시스템 구조

MIRAGE 시스템에서 사용하는 가상문서는 문서내에 실제 내용을 가지고 있지 않고 링크를 이용하여 어떤 문서의 전체 또는 일부분을 내포하거나 참조하는 형식으로 구성되어 있다. 링크서버는 가상문서의 링크 정보를 분리 저장 및 관리하는 서버이다. 링크서버가 저장하는 링크정보는 사용자 에이전트로부터 전달된다. 즉, 사용자 에이전트는 새로운 가상문서에 대한 처리를 하여, 이에 대한 링크정보를 추출한다. 사용자 에이전트가 추출한 링크정보를 저장하라는 요청을 링크서버에게 하면 이때 링크서버가 이 요청을 받아들여 링크정보를 저장하는 것이다. 저장된 정보는 검색서버에서 링크 정보 요청이 있을 경우 링크서버를 통하여 데이터베이스에서 링크를 추출하여 검색서버에 전달한다. 그림 3.2 에서는 현재 구현된 부분은 점선으로 이루어진 사각형 안에 있고 기존의 MIRAGE 시스템과 연동을 하기 위해 인터페이스를 통하여 운영이 되어진다.

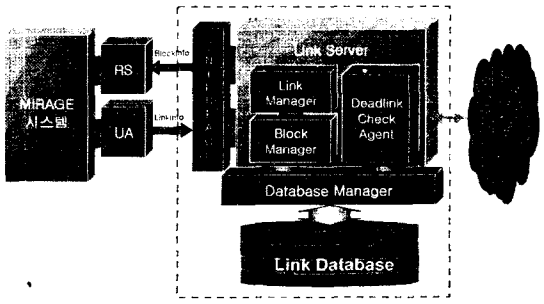

그림 3.2 링크서버의 전체 구조

● 링크 관리자 (Link Manager)

링크는 크게 내포링크와 참조링크로 분류되어 데이터베이스에 저장이 되어지고 관리가 이루어진다. 새로 생성되는 링크를 저장, 변경, 삭제가 가능하다.

● 블록 관리자 (Block Manager)

실제로 운영이 되는 MIRAGE 시스템은 개별 링크정보를 관리하는 것이 아니라, 블록 아이디를 통하여 내포링크와 참조링크를 관리하고 있다.

● 데이터베이스 관리자 (Database Manager)

데이터베이스관리자는 데이터베이스에 관련된 모든 요구사항을 처리할 목적으로 설계가 되어있다.

● 데드링크 검사 에이전트 (Deadlink Check Agent)

웹 상의 문서는 수시로 삭제가 되거나 변경이 이루어진다. 데드링크 검사 에이전트는 주기적으로 링크의 유효성을 검사하고, 문제를 발견하면 오류를 기록하여 다음에 사용 할 경우 오류사항을 알려준다.

3.3. 데이타베이스 설계

링크서버에서 관리하는 데이터베이스의 디자인은 그림 3.3와 같이 구현 되어져 있다. ELINK는 독자적인 테이블로 구성이 되어져 있으며, RLINK는 Destination 테이블과 1:N의 관계를 가지고 있다. 하나의 RLINK를 표현하기 위해서는 하나의 Source와 여러 개의 Destination을 가지게 된다. Source는 RLINK 레코드에 삽입이 되어지고, Destination은 테이블을 따로 두어 저장을 한다. 저장이나 추출을 하기 위해서는 서로의 연관성을 가지고 RLINK 관리를 하게 된다. MIRAGE 시스템에서 처리하는 기본적인 속성에 추가적으로 필요한 속성이 만들어졌다. 추가된 속성은 블록 관리와 링크의 유효성 검사가 가능하도록 디자인이 되어졌다. 가장 중요한 변화는 각 링크에 각기 고유의 ID를 가지고 있어서 서로의 데이터를 연결하기 위한 key로 사용되고 있다.

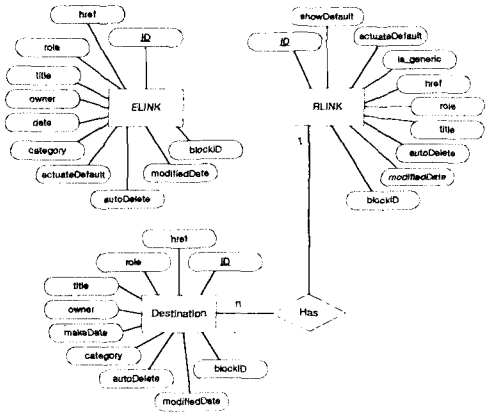

그림 3.3 링크 데이터베이스에 대한 E-R Diagram

3.4. 데드링크 검사 원리

데드링크를 검사하기 위해서는 크게 2가지 메소드로 설명이 되어질 수 있다. 그림 3.4는 DeadlinkDetection 메소드와 isValid 메소드에 대해 설명하고 있다. DeadlinkDetection 메소드는 링크 테이블에서 모든 링크 레코드를 읽어서 검사하는 메소드이다. isValid 메소드는 데드링크를 검출하는 알고리즘의 핵심으로 Shutdown, Update 그리고 Delete를 검출하는 메소드이다.

```
DeadlinkDetection()
  Let L denote a link table

  for each record l ∈ L
    let Γ(l) denote the URI
    If isValid(Γ(l)) = false then
      set l's state to invalid
  end
end
```

```
isValid(σ)
  σ : a URI string
  if σ is shutdown then
      return false
  else If σ is deleted
      return false
  else If σ is modified then
      notify to the system
      return false
  return true
end
```

그림 3.4 데드 링크 검사 알고리즘

4. 응용 프로그램

링크작업은 4.1과 같은 응용 프로그램 도구를 이용하여 내포링크, 참조링크를 추가, 삭제, 변경이 가능하다. 관리자는 원하는 레코드로 이동을 하여 쉽게 원하는 레코드를 관리 할 수 있다. 또한 블록 관리 프로그램과 데드링크 검사 에이전트 프로그램이 존재한다.

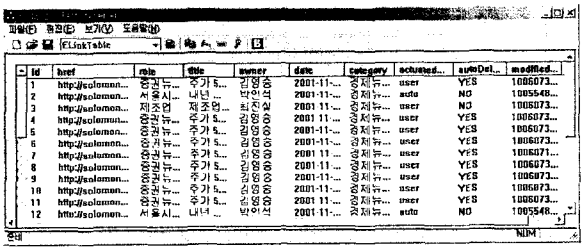

그림 4.1 링크 관리 응용 프로그램

5. 결론

설계한 링크서버는 MIRAGE 시스템에서 사용하는 링크 정보를 보다 효율적으로 저장, 변경, 삭제 그리고 검색의 기능을 제공한다. 링크정보를 관리하기 위해 링크 데이터베이스를 관리하고 있고, 인터페이스를 통해 MIRAGE 시스템은 쉽게 링크정보를 관리할 수 있다. 데드 링크 검사 에이전트는 독립적으로 데이터베이스에 저장된 링크의 유효성 검사를 주기적으로 검사하고, 유효하지 않은 링크에 대해서는 데이터베이스에 오류정보를 기록하게 되어 링크 정보의 신뢰성을 향상시키도록 설계가 이루어졌다.

6. 참고문헌

[1] A. Heuer, H. Meyer, B. Porst and P. Titzler, BlueView: Virtual Document Servers for Digital Libraries, Proceedings of the IEEE on Advances in Digital Libraries, pages 207-217, 2000.

[2] C. Jayawardana, K. P. Hewagamage and M. Hirakawa, Virtual Authoring based on the Shallow Copy Technique for a Collection of Digital Documents, Proceedings of the International IEEE Symposium on Multimedia Software Engineering, pages 77-84, 2000.

[3] H. Kikuchi, Y. Mishina, M. Ashizawa, N. Yamazaki, and H. Fujisawa, User Interface for a Digital Library to Support Construction of a "Virtual Personal Library", Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems, pages 429-432, 1996.

[4] S.H. Myaeng, M.H. Lee, J.H. Kang, E.I. Cho, Y.B. Lee, D.S. Lim, J.M. Lim, H.J. Oh, and J.S. Yang, A Digital Library System for Easy Creation/Manipulation of New Documents from Existing Resource, RIAO, pages 196-207, 2000.