MIRAGE-Ⅲ 디지털도서관에서 가상문서 검색 서버의 설계 및 구현

(Design and Implementation of a Retrieval Server for Virtual Documents in the MIRAGE-Ⅲ Digital Library)

이 용 배 † 맹 성 현 ††
(Yong-Bae Lee) (Sung Hyon Myaeng)

요 약 인터넷이용의 급증에 따라 지식정보화사회 전반의 작업들이 분산환경의 디지털도서관에 저장되어 있는 멀티미디어 컨텐츠를 쉽고 신속하게 찾아 새로운 정보를 생성 또는 공유하는 작업을 통해 이루어진다. 이를 위해 핵심적으로 연구해야 할 부분은 원격지에 흩어져 있는 정보를 효과적으로 연결시켜서 의미있고 일관된 형태로 보여지도록 하는 것이다. 이 목적을 달성하기 위해 디지털도서관에서는 물리적으로 흩어져 있는 정보들이 논리적으로 일관되게 표현될 수 있는 가상공간을 제공해야 하며 가상공간에서 정보가 표현되었을 때 사용자가 원하는 정보를 신속하게 수집하여 제시할 수 있어야 한다. 가상문서(Virtual Document)란 특정 저장장소에 저장되어 존재하던 물리적 문서(Physical Document)들을 대상으로 사용자에게 필요한 일부분 또는 전체를 동적으로 연결하여 통합한 문서를 의미한다.

MIRAGE-Ⅲ 디지털도서관에서는 일반 텍스트문서와 XML로 기술된 구조화된 가상문서를 대상으로 내용기반 검색을 수행한다. 이 시스템에서는 XML 문서구조를 기반으로 부분문서의 검색이 가능하며 문서의 속성 및 계층구조에 대한 검색과 문서간의 링크관계를 이용한 검색도 가능하다. 본 논문에서는 MIRAGE-Ⅲ 디지털도서관에서의 질의어처리기 및 검색기를 설계하고 구현한 방법론에 대하여 기술한다.

키워드 : 가상문서, 디지털 도서관, 구조화된 문서검색, 질의어처리, XML

Abstract One of the most important functions digital libraries need to offer is to help users find necessary information in a distributed environment in the most efficient and effective manner. In order to meet the goal, it is desirable to link scattered pieces of information and present them as a logically coherent whole when the user wants it, so that he or she doesn't need to know their physical location. The virtual document is an integrated document that the total or part of the physical documents stored in a specific repository are linked dynamically.

Our MIRAGE-Ⅲ digital library system provides a content-based retrieval of physical documents and the virtual documents in XML. This system provides a retrieval of partial documents, attributes and hierarchical structures and linked-documents based in structured documents like XML or SGML. In this paper we describe a methodology of design and implementation of the query processor and retrieval server in the MIRAGE-Ⅲ digital library system.

Key words : Virtual Document, Digital Library, Structured Document Retrieval, Query Processing, XML

1. 서 론

디지털도서관의 유용성이 사용자가 분산환경에서 저장되어 있는 정보를 쉽고 신속하게 찾을 수 있도록 도와줄 수 있는지에 달려있으므로 이를 위해 핵심적으로 연구해야 할 부분은 원격지에 흩어져 있는 정보를 효과적으로 연결시켜서 의미있고 일관된 형태로 보여지도록 하는 것

· 본 연구는 한국과학재단지정 소프트웨어 연구센터의 기본 프로그램 연구비 지원에 의해서 수행된 연구결과임.
† 비 회 원 : 충남대학교 컴퓨터과학과
yblee@cs.cnu.ac.kr
†† 종신회원 : 충남대학교 정보통신공학부 교수
shmyaeng@cs.cnu.ac.kr
논문접수 : 2001년 3월 20일
심사완료 : 2001년 12월 3일

이다. 이 목적을 달성하기 위해 디지털도서관에서는 물리적으로 흩어져 있는 정보들이 논리적으로 일관되게 표현될 수 있는 가상공간을 제공해야 하며 가상공간에서 정보가 표현되었을 때 사용자가 원하는 정보를 신속하게 수집하여 제시할 수 있어야 한다. 가상문서[1]란 특정 저장장소에 저장되어 존재하던 물리적 문서들의 일부분 또는 전체를 연결하여 구성하고 재현시에 이들을 동적으로 통합하여 브라우징 할 수 있는 문서를 의미한다. 새롭게 생성된 가상문서에는 실제 데이타는 존재하지 않고 기존에 존재하던 컨텐츠로의 링크들만 갖게 된다.

가상문서 기반 디지털도서관은 단순문서의 저장 및 검색 기능을 가지고 있는 전통적인 도서관 개념을 확장하여 사용자가 기존의 분산환경에 존재하던 디지털문서들을 재사용하여 새로운 문서를 생성하고 저장할 수 있다. 가상문서의 링크가 사용자에게 보여질 때 목적 컨텐츠가 문서 안으로 삽입되어 내용이 보여지는 내포링크와 문서 안에서 다른 문서로 항해할 수 있는 앵커역할을 하는 참조링크로 구성되므로 이러한 특성을 가진 가상문서를 검색하기 위한 디지털도서관에서의 검색서버는 내용을 기반으로 한 내포링크와 참조링크의 검색이 기본적으로 이루어져야 한다. 즉, 가상문서가 문서 재현시에 동적으로 구성되므로 이를 검색하기 위한 새로운 색인 방법과 검색 방법이 필요하며 이를 MIRAGE-Ⅲ 디지털도서관 시스템의 검색서버에서 지원할 수 있도록 설계하였다.

본 논문의 연구과정에서 개발된 MIRAGE-Ⅲ 디지털도서관 시스템에서는 가상문서라는 새로운 개념의 문서를 정의[1]하였고 XML(eXtensible Markup Language)을 이용하여 가상문서를 기술하였다. 기존의 SGML (Standard Generalized Markup Language) 또는 XML과 같은 언어로 기술된 구조화된 문서를 검색하는데 필요한 기능[2,3,4,5,6]은 내용검색, 구조검색, 속성검색이 있었지만 MIRAGE-Ⅲ의 검색서버에서는 기존의 검색기능 이외에 XML로 기술된 가상문서에 대하여 메타데이타 검색을 위한 속성검색 및 기존의 구조화된 문서검색에서 지원하지 않았던 문서간의 링크관계를 이용한 내용검색을 지원할 수 있도록 색인기와 검색기를 설계한 것이 특징이다.

또한, XML과 같은 구조화된 문서 검색시 사용자들은 내용검색, 구조검색, 링크검색, 속성검색을 각각 독립적으로 요구하지 않고 각 검색을 혼합한 복합질의어로 검색할 수 있는데 이러한 기능을 지원할 수 있도록 MIRAGE-Ⅲ의 검색서버에서는 가상문서에 적합한 새로운 질의모델을 정립하고 질의어처리기를 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서 현재까지의 구조화된 문서검색에 대한 국내외 연구동향을 기술하고 3장에서는 MIRAGE-Ⅲ의 특징에 관하여 설명하며 4장에서는 가상문서 검색서버의 구성 및 설계에 관하여 기술한다. 5장에서는 가상문서 색인방법과 6장에서는 질의어처리 설계에 관하여 상세히 설명한다. 7장에서는 검색서버를 실험한 내용을 기술했으며 마지막 8장에서는 결론 및 향후 연구과제에 관하여 기술한다.

2. 관련연구

기존의 정보검색 시스템에서는 검색단의를 문서로 하여 키워드 검색을 하면 문서목록을 결과로 가져오며 사용자의 요구시에 전문을 제시하는 것이 일반적인 형태이다. 그러나 대부분의 문서들은 구조를 가지고 있는데, 논문의 경우에 제목, 장, 절 등의 논리적 구조를 가지고 있으며 신문의 경우에는 날짜, 신문명, 섹션, 기사, 기자명, 광고 등의 구조를 취하고 있다. 이러한 문서의 논리적 구조를 검색에 이용하면 기존의 전문검색 방법보다는 문서의 논리적 구조에 근거한 가중치 검색과 같은 다양한 검색을 할 수 있는 장점[4,7]이 있다.

현재 구조화되어진 문서는 하이퍼미디어 형태로 표현될 수 있는데 이를 기술하기 위한 도구로는 디지털 문서교환을 위한 문서기술 표준인 SGML[8,9]과 SGML에 시간적 공간적 개념을 확장한 하이퍼미디어 기술언어인 HyTime(Hypermedia/Time-based Structuring Language)[10], SGML을 웹에서 응용한 HTML (Hypertext Markup Language) 및 HTML의 차세대 버전인 XML[11,12,13,14] 등이 있다. XML은 웹과 함께 인터넷 응용분야가 활성화됨에 따라 전자상거래 분야[15]나 디지털도서관 관련 문서의 검색 및 표현[1,16,17] 등에서 활발한 연구가 진행되고 있다.

XML/SGML과 같은 언어로 기술된 구조화된 문서를 검색하기 위해서는 다음과 같은 작업이 필요하다.

- 내용기반 검색, 구조기반 검색, 속성기반 검색이 수행될 수 있도록 색인구조를 설계
- 사용자가 찾고자 하는 문서를 정확히 찾을 수 있도록 명령하는 질의어를 처리
- 구조화된 문서특성에 맞는 검색모델을 제시하여 효과적인 검색결과를 얻을 수 있게 해주는 작업

구조화된 문서검색을 지원하기 위한 인덱스[2,3,4,6]에는 기존의 하이퍼텍스트 검색에서 진행되어진 항해와 키워드를 이용[18]하여 해당하는 질의어가 문서 내에 포함되는가를 묻는 내용검색 이외에 문서가 어떤 계층구조를 형성하고 있는가에 대한 구조검색과 각 계층단위별로 어떠한 속성값을 내포하느냐에 대한 속성검색을

지원할 수 있는 구조가 필요하다. 또한 XML과 같은 하이퍼미디어 표현 언어로 기술된 문서들은 임의의 엘리먼트에서 다른 엘리먼트로 링크를 구성할 수 있는데 이러한 링크정보에 대한 검색을 수행하기 위해서는 위에서 설명한 색인구조[2,3,4,6] 이외에 링크정보 검색을 위한 색인구조가 추가적으로 필요하다.

질의어 구성[2,3,5,19]은 구조화된 문서의 내부구조를 모르는 정보이용자에게도 투명한 검색결과를 보장하기 위해 사용자가 자연어질의를 입력하였을 때, 이들을 컴퓨터가 인식할 수 있는 정형화된 질의어로 만든 후 이 정형화된 질의어를 구조화된 문서 검색기가 빠르게 검색할 수 있도록 작은 단위의 질의어로 재구성할 필요가 있다. 현재까지의 질의연구에서는 자연어질의를 정형화된 질의어로 바꾸는 연구보다는 정형화된 질의어를 작은 단위로 잘라 재구성[20]하거나 정형화된 질의어의 설계[21]에 대한 연구가 진행되어 왔다. 구조화된 문서 검색을 위한 색인구조나 질의어처리의 전반적인 부분에서는 Tuong의 연구[2,3,22]에서 내용기반 검색, 구조기반 검색, 속성기반 검색을 지원하는 인덱스 구조와 이를 검색하기 위한 질의어(SCL) 개발에 관한 방법론을 소개하였고, 특히 Tuong은 속성 기반 검색을 위한 인덱스를 구조인덱스와 함께 이용하므로 색인에 따른 저장공간을 최소화하였다.

새로운 검색모델로 SGML 문서를 내용기반으로 빠르게 검색할 수 있는 연구[4]가 수행되었는데, 여기서는 추론망모델이 근간이 되었고 검색의 효율성을 위해 새로운 저장구조를 구현하였다. 이 연구에서 제안한 모델은 SGML 엘리먼트 간의 다양한 구조적 관계를 질의로 표현할 수 있을 뿐 아니라, 문서를 검색할 때도 엘리먼트 단위 검색을 한 후 통합하는 방법을 적용하여 일반문서 검색 방법보다 검색 신뢰도를 향상시킬 수 있음을 보였다.

MIRAGE-III 디지털도서관에서는 가상문서라는 새로운 개념의 문서를 정의하였으며 검색서버에서 이를 검색하기 위한 새로운 질의모델을 정립하고 질의처리기를 설계하였다. 또한 기존의 구조화된 문서검색 기술에서 지원하지 못했던 문서간의 링크관계를 검색할 수 있도록 색인구조와 검색기를 설계한 것이 특징이다.

3. MIRAGE-III 디지털도서관의 특징

3.1 가상문서

현재 정의된 가상문서에서 사용되는 링크들은 다음과 같이 분류될 수 있다[1].

- 내포링크(embedding link)와 참조링크(referential link)

내포링크와 참조링크는 가상문서에서 링크를 생성할 때 사용목적에 따라 분류한 것이다. 내포링크는 가상문서 재현시에 링크의 목적 컨텐츠가 직접 문서 안으로 삽입되어 나타나는 링크를 의미하고, 참조링크는 컨텐트가 문서 안으로 삽입되지 않고 앵커로만 남아 사용자의 선택시에 항해나 브라우징할 수 있도록 표시하는 링크를 의미한다. 예를 들어, 내포링크는 웹 문서의 이미지 삽입과 유사한 기능을 지원하고 참조링크는 하이퍼링크와 유사한 개념이다.

- 일대일(one-to-one) 대응, 일대다(one-to-many) 대응, 다대일(many-to-one) 대응 링크

일대일 대응, 일대다 대응, 다대일 대응 링크는 링크의 대응관계에 의한 분류로 일대일 대응은 링크의 목적 컨텐트가 유일한 것을 의미하며, 일대다 대응은 링크의 목적 컨텐트가 두개 이상인 링크를 의미하며, 다대일 대응은 여러 개의 링크가 같은 하나의 목적 컨텐트를 갖는 것을 의미한다.

- 특정링크(specific link)와 총칭링크(generic link)

특정링크와 총칭링크는 링크의 시작점에서 보는 관점으로 가상문서의 특정위치에 있는 멀티미디어 개체(단어, 이미지, 소리 등)가 특정 컨텐트를 가리킬 경우, 이를 특정(일대일 대응)링크라 하며, 임의의 도메인에 포함된 모든 특정개체들이 모두 하나의 특정 컨텐트를 가리키는 경우에는 총칭링크(다대일 대응)라 한다.

- 전체링크(total link)와 부분링크(partial link)

전체링크와 부분링크는 링크의 목적지에서 보는 관점으로 전체링크는 링크의 목적 컨텐트가 컨텐트 전체인 것을 의미하고 부분링크는 링크의 목적 컨텐트가 컨텐트내의 일부분인 것을 의미한다.

가상문서의 구성은 가상문서의 틀을 설명하는 허브(Hub)와 가상문서별 출력포멧을 기술한 스타일시트(Style Sheet)로 이루어져 있으며 허브는 다시 참조링크 리스트, 내포링크 리스트, 가상문서의 메타데이타로 구성된다[1]. 내포링크 리스트는 가상문서 안에 내포되는 링크들의 집합으로 구성되며 참조링크 리스트는 가상문서 안에 직접적으로 삽입되지는 않으나 항해를 할 수 있도록 하이퍼링크의 집합들로 구성된다. 메타데이타는 가상문서가 개별적으로 갖는 메타정보들로 구성되는데 본 연구의 디지탈도서관에서 사용한 메타데이타는 더블린코어(Dublin Core) 메타데이타[23,24]의 15가지 속성을 모두 수용하여 구성하였으며, 메타데이타를 포함한 가상문서를 기술하기 위한 도구로는 현재 웹 문서의 표준으로 위치하고 있으며 구조화된 문서를 표현하기에

적절한 XML[11,12,13,14]을 이용하여 기술하였다[1].

3.2 MIRAGE-III 디지털도서관의 구조

[그림 1]에서와 같이 MIRAGE-Ⅲ는 용도에 따라 개인 디지탈도서관(MIRAGE-Lite)과 공용 디지털도서관(MIRAGE-Regular)으로 구분된다. 각각의 전반적인 구조는 같지만 개인 디지털도서관은 개개인이 사용하기 편리하도록 사용자에 의존적으로 설계되었고 공용 디지털도서관은 개인 디지털도서관보다 사용자에이전트 기능을 확장시켜 같은 MIRAGE계열의 디지털도서관 시스템이나 분산환경하의 다른 디지털도서관 시스템들과 서비스를 교환할 수 있으며 메타검색을 할 수 있는 기능이 추가된다.

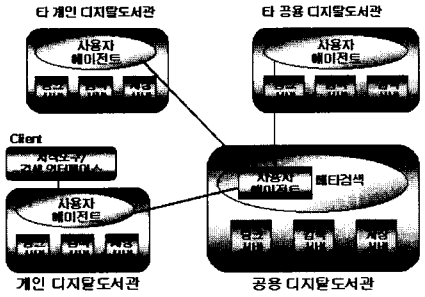

그림 1 분산환경의 MIRAGE-Ⅲ 디지털도서관

다음은 개인 디지털도서관과 공용 디지털도서관 시스템의 공통 모듈에 대한 간략한 설명이다[1].

- 사용자 에이전트

사용자 에이전트는 사용자가 문서저장이나 검색, 삭제 등을 요구할 경우에 서비스를 분류하여 관련 서버에 전달한다. 문서삽입이나 삭제의 경우에는 해당 문서정보를 검색서버나 링크서버, 저장서버에게 전달하여 각각의 서버에서 이 문서정보를 이용하여 문서를 삽입하거나 삭제할 수 있도록 한다. 문서검색을 요구할 경우에는 정형 복합질의어를 검색서버에 전달하며 검색기로부터 검색된 엘리먼트ID(이하 블록ID)를 인터넷상에서 디지털자원의 유일한 식별자(Uniform Resource Identifier, 이하 URI)로 변환하여 사용자에게 검색결과를 전달한다. 사용자가 URI에 매핑되는 문서를 재현할 경우에는 저장서버로부터 URI에 해당하는 문서를 받아와서 사용자가 볼 수 있는 형태로 변환하여 전달한다.

- 검색서버

검색서버는 사용자 에이전트로부터 질의어를 받아 저장서버가 저장하고 있는 블록의 조합으로 이루어진 가상문서와 일반 물리적 문서를 대상으로 검색을 수행하는데, 가상문서와 물리적 문서의 검색은 문서의 특성상 블록 단위의 검색을 할 수 있다. 정확한 문서 검색을 위해 문서 단위의 메타데이타를 검색할 수 있으며 메타데이타 검색은 검색결과를 여과해주는 기능을 한다. 또한 사용자의 링크검색요구를 수행하기 위해 문서나 블록에서 나가고 들어오는 링크정보들을 링크서버로부터 받아와서 처리한다.

검색서버는 사용자 질의어에 대한 검색결과로 블록단위의 가상문서와 문서단위의 일반물리문서, 가상문서를 같이 검색하므로, 서로 다른 문서 컬렉션으로부터의 결과를 하나로 합하여 다시 재랭킹한 다음 사용자에게 전달한다.

- 링크서버

가상문서는 문서내에 실제내용을 가지고 있지 않고 링크를 이용하여 어떤 문서의 전체 또는 일부분을 내포하거나 참조하는 형식으로 구성되어 있다. 링크서버는 검색서버의 링크정보 요청시에 블록이나 문서에서 나가고 들어오는 링크정보를 검색하여 검색서버에 전달한다. 또한 문서내의 링크정보를 따로 저장 관리하므로 검색의 효율을 높이는 기능을 한다.

- 저장서버

저장서버안에 저장된 문서들은 텍스트, 비디오, 오디오, 이미지 등의 문서 혹은 가상문서 형태이며, 각각의 문서들은 모두 URI를 갖는다. 저장서버의 주요기능은 일반문서나 가상문서를 저장 관리하는 일이다. 즉, 사용자 에이전트로부터 문서삽입 또는 삭제 요구를 받아 해당문서를 삽입 또는 삭제하며, 사용자가 일반문서나 가상문서를 재현할 경우에는 URI에 해당하는 문서를 찾아 사용자 에이전트에 전달하는 기능을 수행한다.

- 저작도구

디지털도서관 사용자들은 저작도구를 이용하여 타 문서의 일부 또는 전체에 링크를 생성하여 XML로 기술된 가상문서를 만들 수 있으며, 이미 만들어진 가상문서나 일반문서 또는 웹 문서를 재현할 수 있다. 또한 생성된 가상문서를 사용자에이전트에게 저장을 요구할 수 있다.

- 검색 인터페이스

검색 인터페이스는 사용자로 하여금 쉽게 디지털도서관에 접속하여 디지털도서관 내의 가상문서나 일반문서를 고려하지 않고 문서를 검색할 수 있도록 도와준다. 사용자의 질의를 검색 인터페이스는 자동으로 컴퓨터가 이해할 수 있는 복합질의로 구성하여 사용자 에이전트에 전달하며 검색된 결과를 재현하는 기능을 수행한다.

4. 검색서버의 구성 및 설계

MIRAGE-Ⅲ 디지털도서관 시스템은 저작도구, 검색

인터페이스, 사용자 에이전트, 링크서버, 저장서버는 JAVA 1.2.2로 검색서버는 GNU C++로 구현되었으며 사용자 에이전트와 각 서버들 사이의 통신은 ORBacus CORBA환경에서 동작한다. 본 논문에서는 특히 검색서버의 설계과정을 중심으로 기술한다.

검색서버는 사용자의 복합질의어를 검색서버가 빠르게 검색할 수 있는 단위질의어로 변환하여 기존의 디지털도서관 시스템에서 지원하던 내용검색, 구조검색 이외에 속성검색이나 링크관계 검색을 수행할 수 있도록 하며 사용자에게는 가상문서와 더불어 기존의 디지털도서관에 존재하던 물리적문서를 구분하지 않고 분산환경에서 투명한 검색결과를 전달하는데 목적을 있다.

MIRAGE-Ⅲ 디지털도서관의 검색서버는 기존의 문서검색에서 수행했던 방식과는 달리 질의어부터 XML로 기술된 가상문서와 멀티미디어 문서검색을 위한 복합질의어가 입력되며 이 복합질의어를 분류하여 메타데이타검색기, 블록검색기, 구조검색기, 링크검색기, 일반문서검색기가 빠르게 검색할 수 있도록 단위질의어로 변환하여 검색을 수행한다.

검색연산시에 수행되는 각 모듈간의 상호작용은 [그림 2]와 같다.

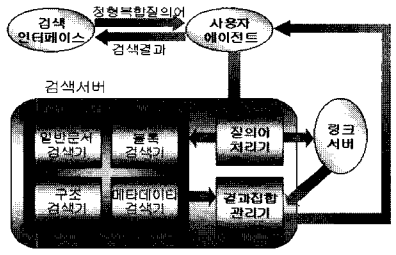

그림 2 검색연산시 서버들간의 관계

여기서 사용자 에이전트가 사용자의 정형복합질의를 확인하고 검색서버로 전달하면 검색서버의 질의어처리기가 정형복합질의어를 일련의 단위질의어로 변환시키고 각 단위질의어를 검색서버로 보낸다. 4개의 검색기와 링크서버가 검색한 후 결과를 조합하여 사용자 에이전트로 보내면, 사용자 에이전트가 검색결과를 사용자에게 보낸다. 검색서버의 각 모듈에 대한 설명은 아래에 기술한다.

- 질의어처리기

질의어처리기는 검색하기 위한 질의어가 내용에 대한 질의와 링크에 대한 질의가 혼합된 복합질의어가 형태로 들어오므로 이를 문서검색기, 블록검색기, 메타데이타검색기, 링크검색기가 각각 수행할 수 있는 단위질의어의 리스트로 바꾸어주는 역할을 한다. 질의어처리기에서 사용자 에이전트로부터 입력으로 받는 복합질의어는 사용자들의 일반적인 검색요구인 자연어질의를 컴퓨터가 처리할 수 있는 형식을 갖춘 질의로 변화시킨 구조적 질의를 뜻한다. 복합질의가 질의어 처리기를 통하여 나오면 검색기가 처리하기 쉬운 일련의 단위질의어로 바뀌는데, 단위질의어란 문서검색기, 블록검색기, 메타데이타검색기, 링크검색기가 인덱스를 한 번이나 두 번 검색하여 처리할 수 있도록 간단하게 바뀐 질의어를 의미한다. 질의어처리기에 대한 상세한 설명은 5장에서 기술한다.

- 메타데이타 검색기

각각의 가상문서에 들어있는 메타데이타는 저자, 제작일, 문서요약내용 등을 포함하는데 XML로 기술될 때 메타데이타는 하나의 엘리먼트를 생성하며 더블린코어의 15가지 속성[17, 18]이 들어간다. 이 속성 값에 대한 검색인 메타데이타검색은 정확한 문서 검색을 위해 검색결과를 여과해 주는 기능을 하며, '홍길동이 만든 문서들을 찾아라.', '2000년 1월 1일 이후의 문서를 찾아라.', '가상문서의 요약내용이 영화 메트릭스와 관련된 문서를 찾아라.'와 같이 직접 메타데이타만으로 메타데이타 데이타베이스를 참조하여 가상문서를 검색할 수도 있다.

- 블록검색기

블록이란 임의의 문서일부 또는 전체를 참조 혹은 내포하여 가상문서가 구성될때, 참조 혹은 내포되는 단위를 본 연구에서는 '블록'이라 정의하였다. 가상문서가 XML로 기술될때 블록은 각각의 엘리먼트로 매핑되고 블록단위 검색질의어가 들어오면, 블록검색기는 미리 색인되어 있는 내용인덱스를 참조하여 내용기반의 엘리먼트 검색을 수행한다.

- 구조검색기

구조검색기는 구조화된 문서 검색에서 문서의 계층구조에 대한 검색이다. 구조검색은 구조화된 문서의 특성에 따른 사용자의 요구에 의존적으로 검색이 수행되며 부모(Parent), 자식(Child), 형제(Sibling), 순서(Order), 조상(Ancestor) 등의 연산을 할 수 있어야 한다. 현재 가상문서가 허브밑에 링크들로 구성된 2개의 계층으로만 이루어져 있으므로 부모-자식과의 관계검색이 빈번하며 형제, 순서, 조상 등을 찾는 질의는 가상문서의 특성상 무의미하다.

사용자의 질의가 특정 엘리먼트를 포함한 가상문서를 찾을 때 내용검색과 함께 구조검색을 수행한다. 예를 들어, '타잔이 포함된 문서를 찾아라.'라는 복합질의가 들어오면 먼저 블록검색기기가 '타잔'을 포함한 엘리먼트

검색을 수행한 후, 구조검색기가 검색 결과들을 대상으로 부모관계에 있는 엘리먼트들을 검색해 낸다.

- 일반문서검색기

일반문서는 기존의 분산환경에 존재하던 멀티미디어 문서들을 의미한다. 가상문서는 이 문서들을 이용하여 재결합된 문서이며, 디지털도서관 정보이용자는 가상문서나 일반문서를 구분하지 않고 질의를 할 수 있다. 검색결과의 투명성을 보장하기 위해서는 사용자가 문서검색을 요구할 경우, 가상문서 이외에 일반문서들도 검색하여 결과를 사용자에게 전달해야 한다. 이때 일반문서 검색기가 일반문서 색인을 검색하는 역할을 한다.

- 링크서버

링크서버에서는 블록을 참조하고 있는 블록(이하 Inlink블록)정보들과 블록이 참조하고 있는 블록(이하 Outlin블록)정보들을 저장하고 있으며 검색서버의 링크정보 요구시에 블록에 대한 Inlink블록정보와 Outlink블록정보를 전달해 준다. [그림 3]은 블록간의 참조관계를 갖는 Inlink블록과 Outlink블록을 보여주고 있다.

- 결과집합 관리기

단위질의어에 대한 검색을 모두 수행하면, 결과집합 관리기에는 최종 검색결과가 들어간다. 이 검색결과들을 정렬할 때에는 일반문서집합과 가상문서집합이 서로 다른 특성을 가진 문서집합이라는 것과 링크검색결과의 링크개수에 가중치를 두어 정렬할 수 있다. 현재의 정렬방법은 가상문서의 메타데이타에서 검색된 문서를 상위

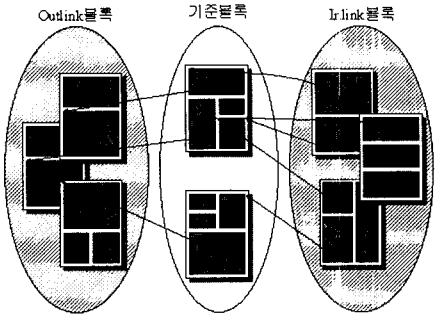

그림 3 블록간의 참조관계

로 위치시키며 가상문서의 내용에 의해 검색된 문서를 그 다음에 위치시키고 일반문서의 검색결과를 마지막 순위로 할당한다.

- 검색 인터페이스

검색 인터페이스는 분산환경의 디지털도서관에 접속된 정보이용자가 디지털도서관 안의 가상문서나 일반문서를 고려하지 않고 문서를 검색할 수 있도록 도와주며, 사용자가 복잡한 가상문서의 검색 질의어를 이해하지 못할지라도 단순한 선택만으로 자동으로 질의어를 만들어 사용자 에이전트에게 전달한다. 또한 사용자 에이전트로부터 검색결과를 받아 디스플레이하고 검색결과로부터 문서가 선택될시에는 스타일이 적용된 XML문서를 재현할 수 있다.

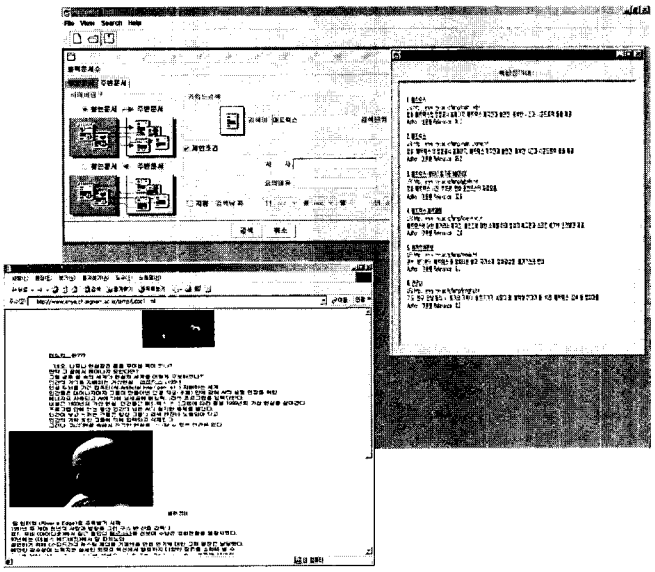

그림 4 MIRAGE-Ⅲ 검색인터페이스

[그림 4]는 가상문서검색을 위한 사용자 인터페이스이다. 왼쪽위의 창은 검색질의를 위한 창으로 검색단추를 누르면 정형복합질의어를 구성하여 사용자 에이전트로 전송한다. 질의창의 왼쪽그림에서 링크조건을 선택할 수 있게 해주며 키워드 입력란에 검색키워드를 입력한 후 메타데이타조건을 선택할 수 있다. 검색명령을 수행한 후 사용자 에이전트로부터 받은 검색결과 리스트는 오른쪽 창에서 보여지며, 검색결과 리스트 중 맨 위의 결과문서를 선택했을 때 왼쪽 아래 창에서 선택된 XML문서가 재현된다.

5. 가상문서 색인기

디지털도서관에서의 중요한 기능은 광대한 양의 문서검색이며, 사용자의 질의에 대한 빠른 검색을 수행하기 위해서는 색인작업이 필수적이다. 색인작업은 기존의 대용량 문서집합을 한번에 색인하던 방법과는 달리, 다중사용자의 문서저장 요구시마다 각 문서에 대한 색인작업이 실시간으로 수행되어야 한다.

특히, 가상문서는 XML로 기술된 구조화된 문서이므로 일반문서의 색인방법과는 달리 추가적인 정보검색을 위한 색인방법이 필요하다. 일반문서의 색인시에 단어별 문서정보추출만을 했던 기존의 방법과는 달리 가상문서는 블록단위의 링크로 구성되므로 단어별 블록정보 추출과 함께 임의의 블록이 참조하는 블록이 무엇인지에 대한 링크정보 추출이 필요하다. 또한 문서단위 검색을 위해 임의의 블록이 어떤 가상문서 안에 포함되었는지에 대한 구조정보 추출이 필요하며 각각의 가상문서에 포함된 메타데이타에 대한 정보추출도 필요하다.

이를 위한 색인구조는 제 4장의 검색서버에서 설명한 내용검색기, 구조검색기, 메타데이타검색기, 링크서버가 각각 검색할 수 있는 내용인덱스, 구조인덱스, 메타데이타인덱스, 링크인덱스로 구성된다.

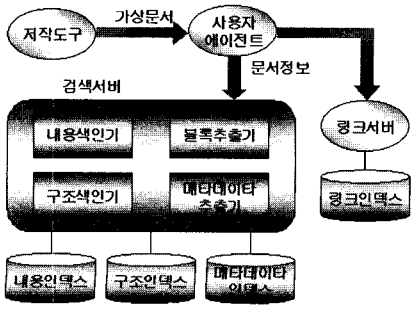

그림 5 색인연산시 시스템 구성도

[그림 5]에서 사용자 에이전트는 사용자의 문서저장 요구시에 가상문서를 처리하여 XML 돔(Document Object Model, DOM)을 구성한 다음 검색서버 및 링크서버가 색인을 위해 필요한 형태로 추출하여 전달해준다. 사용자 에이전트가 가상문서에서 추출하여 검색서버에 전달해주는 문서정보는 [표 1]과 같다.

표 1 색인연산시 문서정보

데이타	형 식
문서정보	{VDocID, (BlockID, LinkType, Content)[], Metadata}
VDocID	가상문서의 식별자로 긴 정수형으로 표현된다.
BlockID	블록의 식별자로 긴 정수형으로 표현된다.
LinkType	가상문서 안에 있는 블록이 내포링크인지 참조링크인지를 구별하기 위한 기호로 정수형으로 표현된다.
Content	각 블록안의 실제 내용으로 스트링형으로 표현
Metadata	메타데이타는 더블린코어의 메타데이타 속성을 모두 사용하며 현재는 저자, 날짜, 문서요약만을 메타데이타에 넣는다. 즉, 메타데이타는 저자,날짜, 요약으로 구성되며 저자는 스트링형, 날짜는 긴 정수형, 요약은 스트링형으로 표현된다.
문서내용	문서정보에서 VdocID만을 제외한 나머지부분으로 {(BlockID, LinkType, Content)[], Metadata}가 해당된다.

위의 데이타를 이용하여 검색서버의 색인연산시에 동작하는 모듈들의 기능은 다음과 같다.

- 블록추출기

블록추출기는 사용자 에이전트로부터 받은 문서정보를 VDocID와 문서내용으로 분류하고 문서내용에서는 메타데이타를 잘라 메타데이타 추출기로 전달한다. 또한 VDocID는 내용색인기와 메타데이타 추출기에서 문서식별자로 사용할 수 있도록 내용색인기와 메타데이타 추출기로 전달한다.

- 내용색인기

내용색인기는 블록추출기가 분류한 메타데이타를 제외한 문서내용{(BlockID, LinkType, Content)[]}에서 블록단위로 텍스트에 대한 색인작업을 수행하여 내용인덱스를 구성하며, 메타데이타의 속성중 문서요약 부분도 색인하여 따로 색인파일을 구성하지 않고 내용인덱스에 추가시켜 사용하므로 색인구조 저장을 위한 공간을 줄여준다. [그림 6]은 내용인덱스의 구조를 보여준다.

- 메타데이타추출기

메타데이타추출기는 블록추출기가 분류한 메타데이타 중에서 문서요약정보는 내용색인기로 보내고 저자와 날

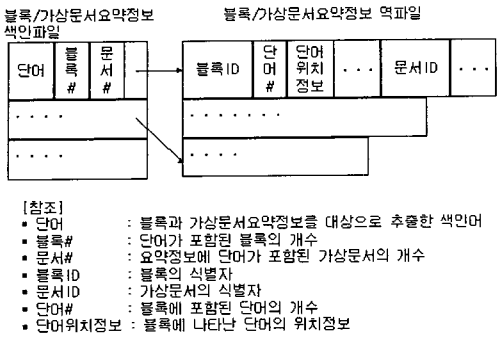

그림 6 내용인덱스의 구조

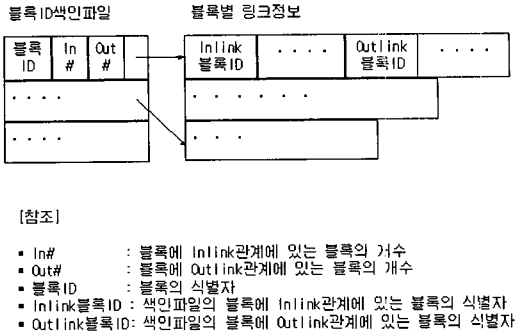

그림 7 링크인덱스의 구조

짜부분을 분류하여 저자별로 날짜별로 검색 가능하도록 날짜인덱스, 저자인덱스에 저장시킨다.

- 구조색인기

구조화된 문서검색에서 구조색인기의 역할은 문서의 특성과 사용자의 검색 요구에 의존하여 엘리먼트 단위로 부모, 자식, 형제, 순서, 조상 등의 구조정보를 뽑아내어 인덱스를 구성하는 일이다. 가상문서 검색은 문서의 특성상 부모-자식과의 관계만 필요하므로 구조색인기는 문서에서 부모-자식 정보만을 추출하여 인덱스를 구성한다. 즉, 가상문서가 블록이라고 부르는 엘리먼트 단위로 색인되어 있으므로 엘리먼트가 아닌 문서전체를 검색할 경우에는 블록을 포함한 가상문서를 찾아줄 수 있도록 가상문서내에 내포링크로 포함된 블록ID와 가상문서ID를 매핑시켜주는 구조를 형성한다.

- 링크색인기

링크서버안의 링크색인기는 사용자 에이전트로부터 링크 색인정보를 입력받아 가상문서의 블록별로 Inlink 블록과 Outlink블록에 대한 색인을 수행하여 링크인덱스를 구성한다. 링크서버는 검색서버의 링크정보 요구시에 링크인덱스를 검색하여 결과를 검색서버에 전달한다. [그림 7]은 링크정보를 색인한 결과인 링크인덱스의 구조를 보여준다.

설계된 색인구조를 기반으로 실제 색인연산은 가상문서가 저장되는 시점에서 실시간으로 작동한다. 클라이언트의 저작도구에서 사용자에이전트로 문서저장을 요구하면 사용자 에이전트에서는 내용인덱스, 구조인덱스, 메타데이타인덱스, 링크인덱스로 색인을 위해 각 색인기에서 필요한 정보들을 구성하여 전송한다. 먼저 메타데이타추출기는 사용자에이전트에서 받은 메타데이타를 대상으로 날짜와 저자를 기준으로 문서를 빠르게 찾아가는 구조를 구성한다. 메타데이타인덱스 구성이 끝나면 구조색인기는 블록식별자를 기준으로 블록을 포함한 가상문서들을 찾을 수 있는 구조를 구성하며 그 후에 내용색인기가 기존의 문서색인 방법과 유사하게 가상문서에 내포되는 블록단위로 추출된 용어를 기준으로 블록을 찾아갈 수 있는 구조를 형성한다. 마지막으로 링크서버에서는 블록단위로 들어오고 나가는 블록들을 빠르게 검색할 수 있는 구조를 구성하면서 하나의 문서 삽입시 색인연산이 종료된다.

6. 질의어 처리기

가상문서는 본 연구에서 새롭게 정의[1]한 문서이고 XML을 이용하여 기술하였다. 이 문서를 검색하기 위해서는 XML문서의 특성과 가상문서 검색 요구조건을 분석하여 이에 적합한 새로운 질의모델이 필요하며 질의모델에 의해 구성된 질의어를 검색기가 이해하여 빠르게 검색할 수 있도록 변환시키는 질의어처리기가 필요하다.

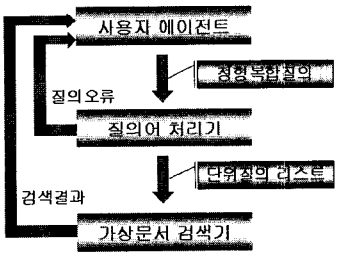

그림 8 질의어 처리기

디지털도서관 사용자들은 가상문서에 대하여 내용검색, 구조검색, 링크검색, 메타데이타검색을 각각 독립적으로 요구하지 않고 각 검색을 혼합하여 한번에 요구할 수 있도록 질의어를 구성할 수 있는데, 이러한 질의어

를 복합질의어라 한다. [그림 8]의 질의어 처리기에서 입력으로 받은 정형복합질의란 사용자들의 일반적인 검색요구인 자연어질의를 컴퓨터가 처리할 수 있는 형식을 갖춘 질의로 변화시킨 구조적 질의를 뜻한다. 정형복합질의가 질의어 처리기에 의해 처리되어 나오면 가상문서 검색기가 처리할 수 있는 일련의 단위질의로 바뀌는데 단위질의란 검색기가 인덱스를 한 번 이나 두 번 검색하여 처리할 수 있도록 간단하게 바뀐 질의어를 의미한다.

6.1 정형복합질의어의 구조

정형복합질의어의 문법규칙은 [표 2]에 기술되어 있다. 가상문서 질의를 위한 키워드는 검색단위지정, 일반기호, 내용검색명령어, 링크검색명령어, 메타데이타검색명령어의 크게 다섯 가지로 구분할 수 있다. 정형복합질의어의 구성은 위 다섯가지 키워드의 조합으로 이루어지며 키워드를 이용한 질의어구성과 이에 대한 세부설명은 [표 3]에서 설명되어 있다.

6.2 단위질의어의 구조

정형복합질의어가 질의어처리기에 의해 처리되어 나온 일련의 질의를 단위질의라 하는데 정형복합질의어를 세분화시켜 블록검색기, 메타데이타검색기, 링크검색기, 구조검색기, 일반문서검색기가 색인구조를 한번이나 두 번 접근하여 검색할 수 있도록 분류해 놓은 단순질의어를 의미한다. 검색서버에서 처리할 수 있는 단위질의어로는 블록검색어, 일반문서검색어, 부모검색어, 저자검색어, 날짜검색어, 요약정보검색어, Inlink검색어, Outlink

표 2 정형복합질의어의 문법

```
CompositeQuery  := Query | Query 'REF' Query | Query 'REFED' Query
Query           := BlockQuery | DocQuery | '{'BlockQuery'}' | '{'DocQuery'}'
BlockQuery      := 'BLOCKS' | 'BLOCKS' ContentQuery
DocQuery        := 'DOCS' | 'DOCS' [ContentQuery] [MetaQuery]
ContentQuery    := '(' QUOTE termlist QUOTE ')'
MetaQuery       := '[' AuthorQ DateQ DescriptQ ']'
AuthorQ         := 'AUTHOR' EQ QUOTE termlist QUOTE
DateQ           := 'DATE' EQ QUOTE DatePeriod QUOTE
DescriptQ       := 'DESC' EQ QUOTE termlist QUOTE
Termlist        := term | term termlist
DatePeriod      := DateForm | DateForm '~' DateForm | DateForm '~' | '~' DateForm
DateForm        := yyyy '/' mm '/' dd
Term            := [0-9|a-z|A-Z|가-힣]+
EQ              := '='
QUOTE           := '\''
```

표 3 정형복합질의어의 키워드 분류 및 사용예

구분	키워드	설명	정형복합질의 사용예
검색단위 지정	blocks, docs	검색단위를 블록 혹은 문서로 지정한다.	blocks('바하 헨델') -> '바하 헨델'을 포함한 블록을 검색한다. Docs('이승엽 홈런') -> '이승엽 홈런'을 포함한 일반문서와 가상문서를 검색한다.
일반기호	'', *, =, ~, {}	내용검색,링크검색,메타데이타 검색에 필요한 기호	{Docs [author=* date='1999/12/25' desc=*]} -> 가상문서를 만든 저자나 내용에 관계없이 1999년12월25일에 만들어진 문서만을 검색한다.
내용검색 명령어	(' ')	어떤 단어가 포함된 문서나 블록을 검색할 경우, 단어를 지정하기 위한 명령어	blocks('하이퍼미디어') -> '하이퍼미디어'를 포함한 블록을 검색한다.
링크검색 명령어	ref, refed	문서나 블록이 참조하거나 다른 문서나 블록에 의해 참조되는 링크정보를 검색한다.	docs ref {blocks ('시네마천국')} -> '시네마천국'을 포함한 블록을 참조하는 모든 문서를 검색한다. blocks refed docs('SGML XML') -> 'SGML XML'을 포함한 문서에 의해 참조되어지는 모든 블록을 검색한다.
메타데이타검색 명령어	[author=' ' date=' ' desc= ' ']	메타데이타를 검색하기 위해 문서의 저자, 날짜, 요약내용을 검색한다.	Docs[author='홍길동'date=* desc='쇼팽 바하 헨델'] -> 문서작성일에 관계없이 저자가 '홍길동'이고, '쇼팽 바하 헨델'의 내용을 포함한 문서를 검색한다.

검색어의 8가지 종류가 있다. 단위질의어의 종류와 형식은 [표 4]에서 설명한다.

표 4 단위질의어의 종류 및 사용예

종류	형식	설명
블록 검색어	blocks('정보검색')	'정보검색'을 포함한 가상문서의 블록을 검색
일반문서 검색어	pdocs('자연어 처리')	'자연어 처리'를 포함한 일반문서를 검색
구조 검색어	parent(result4)	4번째 검색결과집합의 블록을 포함한 가상문서를 검색
저자 검색어	meta[author='현진건']	저자가 '현진건'인 가상문서를 검색
날짜 검색어	meta[date='20000101-']	2000년1월1일 이후의 가상문서를 검색
요약정보 검색어	meta[desc='영화 메트릭스']	요약내용에 '영화 메트릭스'를 포함한 가상문서를 검색
Outlink 검색어	outlink(result2)	2번째 결과집합에 대한 outlink 블록을 검색
Inlink 검색어	inlink(result5)	5번째 결과집합에 대한 inlink 블록을 검색

6.3 질의어 처리결과 및 검색방법

디지털도서관의 정보이용자가 '타이타닉을 포함한 문서중에 제임스 카메룬이 만든 영화가 포함된 문서를 참조하고 있는 문서들을 찾아라.'라는 자연어질의를 보내면 아래와 같은 정형복합질의어로 변환될 수 있다.

docs ('타이타닉') ref {docs ('영화') [author='제임스 카메룬' date=* desc=*]}

위와 같은 정형복합질의어를 사용자에이전트에서 받아 질의어처리 후, 가상문서 검색기가 검색을 수행한다. 질의어처리기를 통한 단위질의 리스트는 아래와 같으며 각 단위검색기는 질의어처리 순서대로 내용인덱스, 구조인덱스, 링크인덱스, 메타데이타 인덱스를 대상으로 검색을 수행하여 [표 5]와 같은 검색결과를 얻는다.

7. 실 험

MIRAGE-Ⅲ 검색서버는 분산환경에서 가상문서 저작과 검색을 계몽사 데이타 집합과 한글테스트 컬렉션(Hangul Test Collection, HANTEC)의 해외과학기술동향 18,442건, 한국여성개발원 문서 110건을 대상으로 리눅스서버와 팬티엄Ⅲ PC하에서 실험하여 그 기능을 확인하였다. 또한 가상문서의 특성을 반영한 복합질의어 30개를 작성하여 질의어처리기를 테스트하였으며 그 중 10개를 추출하여 검색시간을 확인한 결과 평균 1.877초를 기록하였다.

가상문서 검색을 위해 사용된 복합질의어와 그에 따른 검색시간은 [표 6]에서 나타낸다.

표 5 질의처리결과를 이용한 검색방법

검색순서	단위질의어	검색방법
결과1	blocks('타이타닉')	'타이타닉'을 포함한 블록을 내용인덱스를 검색하여 결과1에 저장
결과2	parent (결과1)	결과1의 각 블록들을 포함한 가상문서를 구조인덱스를 검색하여 결과2를 구성한 후 결과1을 삭제
결과3	meta[desc='타이타닉']	가상문서 메타데이타의 요약정보에 '타이타닉'이 포함된 문서를 메타데이타 인덱스를 검색하여 결과3 구성
결과4	결과2 or 결과3	결과2와 결과3을 합집합 연산하여 결과4을 생성하고 사용된 결과2, 3은 삭제
결과5	pdocs ('타이타닉')	'타이타닉'이 포함된 문서를 내용인덱스를 검색하여 결과5를 구성
결과6	결과4 or 결과5	결과4와 결과5를 합집합 연산하여 결과6을 구성한 후, 사용된 결과4, 5는 삭제
결과7	blocks('영화')	'영화'를 포함한 블록을 내용인덱스를 검색하여 결과7에 저장
결과8	parent (결과7)	결과7의 각 블록들을 포함한 가상문서를 구조인덱스를 검색하여 결과8을 생성하고 결과7은 삭제
결과9	meta[desc='영화']	가상문서 메타데이타의 요약정보에 '영화'를 포함한 문서를 메타데이타 인덱스를 검색하여 결과9 생성
결과10	결과8 or 결과9	결과8과 결과9를 합집합 연산하여 결과10을 생성한 후, 사용된 결과8, 9는 삭제
결과11	meta[author='제임스 카메룬']	가상문서의 저자가 '제임스 카메룬'인 문서를 메타데이타 인덱스를 검색하여 결과11 생성
결과12	결과10 and 결과11	결과10과 결과11을 교집합 연산하여 결과12을 생성한 후, 사용된 결과10, 11은 삭제
결과13	inlink(결과12)	결과12의 블록들을 대상으로 링크인덱스를 검색하여 결과13을 생성
결과14	결과6 and 결과13	결과6과 결과13을 교집합 연산하여 결과14를 생성한 후, 사용된 결과6, 13은 삭제하고 결과 14를 사용자 에이전트로 전송

표 6 질의어에 따른 검색시간

번호	정형복합질의어	검색시간
1	docs ('사건') ref {docs ('국제문제') [author='이용배' date=* desc=*]}	1.933
2	blocks('여성')	1.802
3	blocks('3.1운동') ref blocks('민족운동')	1.819
4	docs('축구') refed blocks('월드컵 진출')	1.921
5	blocks('영화') refed docs('액션') [author=* date='19990101~' desc=*]	1.843
6	{docs('올챙이') [author=* date='~20001231' desc=*]} ref blocks('과학')	1.846
7	{docs('거북이')}	1.912
8	{docs('올챙이') [author=* date='~20001231' desc=*]} ref {docs('생물학') [author='이용배' date='~20001231' desc=*]}	1.902
9	docs('에너지') [author='이용배' date='~20001231' desc=*] refed {docs('물리') [author= date=* desc='중학교 과정']}	1.901
10	docs ('교통사고') [author='이용배' date=* desc='항공기 추락사고']	1.890
평균검색시간		1.877

[표 6]에서는 가상문서의 검색시간이 복합질의어에 'docs'를 포함한 문서단위로 검색할 경우(1,4,5,6,7,8,910)와 복합질의어에 'blocks'만을 포함한 블록단위로 검색할 경우(2,3)가 검색시간의 차이가 있음을 보이고 있다. 이것은 문서단위로 검색을 하면 내용인덱스와 구조인덱스 및 일반 물리적 문서 인덱스를 모두 검색하므로 시간이 많이 소모되며 블록단위로 검색을 하면 키워드를 포함한 내용인덱스만을 검색하므로 검색시간이 짧아지기 때문이다.

문서단위로 검색할 경우에도 메타데이타 조건이 있는 가상문서만을 검색(5,6,8,9,10)하는 것과 메타데이타 조건이 없는 가상문서와 일반문서를 검색(1,4,7)하는 것이 검색시간이 더 걸리는 것을 볼 수 있다. 또한 링크관계 검색(1,3,4,5,6,8,9)에서는 전체 검색시간이 별로 영향을 주지 않은 것을 알 수 있었다.

8. 결 론

본 논문에서는 새로운 형태의 디지털도서관인 MIRAGE-Ⅲ의 구조적인 틀을 제시하고, 이를 응용한 디지털도서관 모형을 제작하는 과정에서 저작된 멀티미디어 가상문서에 대하여 사용자에게 투명한 검색환경을 제공하는 검색서버의 설계 및 구현에 대하여 기술하였다.

MIRAGE-Ⅲ 디지털도서관 시스템은 사용자 에이전트와 검색서버, 링크서버, 저장서버로 구성된다. 사용자 에이전트는 사용자의 서비스 요구를 분석하여 각 서버에게 서비스 요구를 명령하며 각 서버로부터 서비스 결과를 받아 사용자에게 전달하는 에이전트 역할을 한다. 검색서버는 가상문서를 색인하고 사용자 에이전트의 정형복합질의를 단위질의로 처리하여 내용검색, 링크검색, 구조검색, 메타데이타검색을 수행한 후 결과를 사용자 에이전트에게 전달하며, 링크서버는 가상문서에서 링크정보를 분리하여 저장관리하고 검색서버의 링크정보 요구를 처리하는 기능을 한다. 또한 저장서버는 디지털도서관에 있는 문서들을 저장하고 관리하는 역할을 한다.

본 연구에서 정의한 가상문서는 XML을 사용하여 구체화시켰다. 가상문서는 문서의 틀을 기술하는 허브와 문서의 스타일을 기술하는 스타일시트로 구성되는데 허브는 다시 내포링크, 참조링크, 메타데이타로 구성된다. 내포링크는 가상문서 안으로 직접 삽입되는 링크를 의미하며 참조링크는 현재 웹 문서의 하이퍼링크와 유사하게 다른 문서로 항해할 수 있는 링크를 의미한다. 가상문서의 메타정보를 기술하는 메타데이타는 더블린코어의 메타데이타 속성 15가지를 모두 수용하였다.

MIRAGE-Ⅲ의 검색서버는 사용자의 정형복합질의어를 검색서버가 처리할 수 있는 단위질의로 바꾸어 기존의 내용검색 이외에 구조검색, 링크검색, 메타데이타 검색을 수행할 수 있도록 하였으며 사용자에게 가상문서와 더불어 기존의 디지털도서관에 존재하던 물리적문서를 구분하지 않고 분산환경에서 투명한 검색결과를 사용자에게 전달할 수 있도록 하였다. 특히, 기존의 XML/SGML과 같은 구조화된 문서검색시에 지원하지 못하던 문서간의 링크관계를 이용한 내용기반 검색을 수행할 수 있도록 설계되었다는 것이 특징이다.

앞으로는 저작과 검색대상 문서를 관리가 되고있는 인트라넷상의 데이타베이스로 국한하지 않고 인터넷상의 웹문서로의 확장이 필요하며 현재 설계를 확장중에 있다. 또한 가상문서 색인기의 효율적인 변경 및 삭제 알고리즘이 필요하며 가상문서 검색결과의 순위결정시 링크정보를 이용한 적합한 검색순위 정렬 알고리즘이 필요하다.

참 고 문 헌

[1] Sung Hyon Myaeng, Mann-Ho Lee, Ji-Hoon Kang, Eun-Il Cho, Yong-Bae Lee, Dong-Soo Lim, Jeong-Mook Lim, Hyo-Jung Oh, Jung-Shik Yang, "A Digital Library System for Easy Creation/Manipulation of New Documents from Existing Resources," Proceedings of RIAO 2000, pp196-208, April 2000.

[2] Tuong Dao, "An Indexing Model for Structured

Documents to Support Queries on Content, Structure and Attributes," Proceedings of ADL '98, 1998.
[3] T. Dao, R. Sacks-Davis and J. A. Thom,"An Indexing Scheme for Structured Documents and its Implementation," Proceedings of the 5th International Conference on Database System for Advanced Applications, April 1997.
[4] Sung Hyon Myaeng, Dong-Hyun Jang, Mun-Seok Kim, Zong-Cheol Zhoo, "A Flexible Model for Retrieval of SGML Documents," Proceedings of ACM SIGIR '98, pp138-145, 1998.
[5] Ian A. Macleod, "Storage and Retrieval of Structured Documents," Information Proceeding & Management, Vol.26, No.2, 1990.
[6] Lee, Y. K., Yoo, S. J., Yoon, K. & Berra, P. B., "Index Structure for Structured Documents," in Digital Library '96, 1996.
[7] 맹성현, 주종철, 문서구조화와 정보검색, 정보과학회지 제16권 제8호, 1998.
[8] Brian E. Travis, Dale C. Waldt, The SGML Implementation Guide, Springer, 1995.
[9] Charles F. Goldfarb, The SGML Handbook, Clarendon Press, Oxford, 1990.
[10] W. Eliot Kimber, "What's New and Cool in HyTime," 1997. (available at http://www.isogen.com/papers/newcool.html)
[11] Simon ST.Laurent, XML A Primer, MIS:Press, 1998.
[12] eXtensible Markup Language(XML) version 1.0, recommendation 1998. (available at http://www.w3c.org/XML/)
[13] Eric Miller, "An Introduction to the Resource Description Framework," D-Lib Magazine, May 1998.
[14] W3C, Resource Description Framework(RDF) Schema Specification 1.0, 2000.(available at http://www.w3.org/TR/2000/CR-rdf-schema-20000327/)
[15] Thibadeau, R. et al., "E-Commerce Catalog Construction: An Experiment with Programmable XML for Dynamic Documents," D-lib Magazine, February 1999.
[16] William Y. Arms, Christophe Blanchi, Edward A. Overly, "An Architecture for Information in Digital Libraries," D-lib Magazine, February 1997.
[17] S. Payette, C. Lagoze, "Flexible and Extensible Digital Object and Repository Architecture (FEDORA)," Proceeding of the 2nd European Conference on Digital Libraries, September 1998.
[18] Maristella Agosti, Information Retrieval and Hypertext, in Information Retrieval and Hypertext, Kluwer Academic Publishers, 1996.
[19] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "Structured Queries," in Modern Information Retrieval, Addison Wesley, pp 106-109, 1999.
[20] 맹성현, 장동현, 이용배, 구조화 정보검색 모델 및 알고리즘 개발에 관한 연구, 한국전자통신연구원 위탁과제 최종보고서, 1998.
[21] 이계준, 신동욱, 권택근, "XML 문서의 검색을 위한 효율적인 색인기법과 질의언어(TQL)의 설계," 한국정보과학회 가을 학술발표논문집 Vol.26, No.2, 1999.
[22] T. Dao, R. Sacks-Davis and J. A. Thom, "Indexing Structured Text for Queries on Containment Relationships," Proceedings of the 7th Australian Database Conference, Jan. 1996.
[23] Dublin Core Community, Dublin Core Metadata Initiative, recommendation 1999. (available at http://purl.org/DC/documents/)
[24] Baker, T, "Language for Dublin Core," D-lib Magazine, December 1998.
[25] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "Structured Text Retrieval Model," in Modern Information Retrieval, Addison Wesley, pp 61-65, 1999.
[26] Klemens Bohm, Adrian Muller, Eric Neuhold, "Structured Document Handling - a Case for Integrating Database and Information Retrieval," Proceedings of the third International Conference on Information and Knowledge Management, 1994.
[27] 맹성현, 분산환경에서의 멀티미디어 가상문서의 표현 및 검색에 관한 연구, 충남대학교 소프트웨어 연구센터 최종보고서, 1999.

이 용 배
1996년 충남대학교 컴퓨터과학과(학사). 1998년 충남대학교 컴퓨터과학과 대학원(석사). 현재 충남대학교 컴퓨터과학과 대학원(박사과정). 관심분야는 정보검색, 자연어처리, 디지털도서관, 장르분류, 지식관리시스템, 하이퍼미디어시스템

맹 성 현
1983년 미국 캘리포니아 주립대학 학사. 1985년 미국 Southern Methodist University(SMU) 석사. 1987년 미국 Southern Methodist University(SMU) 박사. 1987년 ~ 1988년 미국 Temple University 교수. 1988년 ~ 1994년 미국 Syracuse University 교수. 1994년 ~ 현재 충남대학교 정보통신공학부 교수. 관심분야는 정보검색, 자연어처리, 디지털도서관, 자동요약, 자동분류, 지식관리시스템