공학석사 학위논문

제공자와 사용자의 키워드로 만들어진 검색트리를 이용한 XML 문서 검색시스템의 설계

지도교수 조 우 현

이 논문을 공학석사 학위논문으로 제출함

2001년 8월

부경대학교 대학원

컴퓨터공학과

정 성 주

정성주의 공학석사 학위논문을 인준함

2004 년 8 월 31 일

주 심 공학박사 서 경 룡

위 원 공학박사 신 봉 기

위 원 공학박사 조 우 현

목 차

그 림 목 차

표 목 차

제공자와 사용자의 키워드로 만들어진 검색트리를 이용한 XML 문서 검색시스템의 설계

정 성 주

부 경 대 학 교 대 학 원 컴 퓨 터 공 학 과

요 약

본 논문에서는 데이터를 자유롭게 표현할 수 있는 XML문서에 대한 검색 시스템을 제안한다. XML 문서는 태그를 이용하여 내용을 자유롭게 표현할 수 있다. 본 시스템은 태그와 관련된 내용을 사용자가 입력한 단어나 구와 대응시켜 문서를 사용자 중심으로 표현하여 효율적인 검색을 수행한다. 원격지의 XML 문서는 멀티미디어 데이터를 포함하거나 가리키고 있다고 가정한다. 원격지의 XML 문서를 검색하기 위해서 데이터 제공자가 제시한 검색 트리와 사용자가 입력한 검색어로 구성된 검색 트리를 모두 이용하는 방안을 제안한다. 효과적인 검색을 위해 사용자가 입력한 검색어를 선별하여 사용자 모델로 활용하는 방법과 사용자 모델을 바탕으로 높은 재생산도를 얻기 위해 실제 검색에 이용할 모델을 만드는 방법을 제시한다.

Design of XML Document Retrieval System
using Search trees by Provider and User's Keywords

Sung Ju Jung

Dept. of Computer Eng., Graduate School,
Pukyong National University

Abstract

In this paper, We propose a search system for XML document database that can express data freely. The XML document can express contents using tag. Our system that expresses document for user effectively performs to retrieve XML tag corresponding to user's word or phrase. We assume that a remote XML document include multimedia data or indicate one. We propose a method using both a search tree based on an provider's data and a search tree based on user's input key word to retrieve remote XML document. Our system provide user model with extracting portion of user's key word for effectiveness of searching and real search model based on user model to get high reproduction degree.

1. 서론

인터넷의 보급과 사용자 컴퓨터의 증가로 웹 상의 문서나 멀티미디어 데이터를 편리하게 이용할 수 있게 되었다. 과거에는 전송 선로의 제약때문에 주로 웹 문서의 검색만을 할 수 있었지만, 현재는 웹 문서뿐만 아니라 이미지, 동영상 등의 멀티미디어에 대한 검색도 증가하였다. 인터넷 상에 문서나 멀티미디어 데이터의 증가로 데이터 검색에 대한 중요성이 부각되었다. 기존 검색 사이트에서는 웹 문서에 존재하는 단이나 구를 입력하면 원하는 문서를 검색할 수 있다. 또한 웹 문서를 검색하는 경우 문서내의 단어, 구 등을 이용해 검색할 수 있다. 웹 문서의 경우 기존에는 html태그를 이용한 문서가 주를 이루었지만, 문서간의 호환과 표현 능력의 한계로 현재는 XML 표준이 이용되고 있다. 또한 XML 문서의 경우 태그의 성격과 구성을 직접 조작할 수가 있어 다양한 표현을 할 수 있고, 다른 형식의 문서로 변환할 수도 있다. 또한 전송 방법의 발달로 이미지, 음악, 영화 같은 대용량의 멀티미디어 데이터의 전달이 가능해졌다. 원격지에 존재하는 다양한 정보에 대한 검색 실패와 이로 인해 발생하는 네트워크 부하를 막기 위해 효과적인 검색 방법이 필요하다.

기존의 연구는 영화 데이터에 대한 검색 시스템을 제안하였다. 멀티미디어 데이터는 자체에 단어나 구를 가지지 않기 때문에, 데이터 제공자가 제시한 검색어를 바탕으로 사용자가 검색을 할 수 있다. 그러나 멀티미디어 데이터의 경우, 데이터 제공자가 제시한 검색 정보만으로는 그 데이터를 효과적으로 표현하기가 어렵기 때문에 사용자가 원하는 멀티미디어 데이터를 효과적으로 검색할 수 없다. 데이터 제공자가 제시한 검색정보는 제목, 줄거리, 장르 등으로 한정되어 있다. 또한 멀티미디어 데이터베이스가 원거리에 떨어져 있고 데이터베이스의 종류가 다르다면 통합된 검색을 하기가 어렵게 된다.

본 논문에서는 데이터를 자유롭게 표현할 수 있는 XML문서에 대한 검색 시스템을 제안한다. XML 문서는 태그를 이용하여 내용을 자유롭게 표현할 수 있다. 본 시스템은 태그와 관련된 내용을 사용자가 입력한 단어나 구와 대응시켜 문서를 사용자 중심으로 표현하여 효율적인 검색을 수행한다. 원격지의 XML 문서는 멀티미디어 데이터를 포함하거나 가리키고 있다고 가정한다. 원격지의 XML 문서를 검색하기 위해서 데이터 제공자가 제시한 검색 트리와 사용자가 입력한 검색어로 구성된 검색 트리를 모두 이용하는 방안을 제안한다. 효과적인 검색을 위해 사용자가 입력한 검색어를 선별하여 사용자 모델로 활용하는 방법과 사용자 모델을 바탕으로 높은 재생산도를 얻기 위해 실제 검색에 이용할 모델을 만드는 방법을 제시한다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련된 연구와 기존의 연구를 소개하고 문제점을 제기한다. 3장에서는 시스템 제안과 성능 평가를 한다. 4장에서는 결론을 내린다.

2 관련 연구와 기존의 연구에 대한 소개

2.1 관련 연구

웹상에서 문서 서비스를 제공하기 위해서는 서비스의 방법과 문서 표준, 데이터베이스에 대한 지식이 필요하다.[9]

◉XML

XML은 1996년 W3C(World Wide Web Consortium)에서 제안한 것으로서, 웹 상에서 구조화된 문서를 전송 가능하도록 설계된 표준화된 텍스트 형식이다. 이는 인터넷에서 기존에 사용하던 HTML의 한계를 극복하고 SGML의 복잡함을 해결하는 방안으로써 HTML에 사용자가 새로운 태그(tag)를 정의할 수 있는 기능이 추가되었다. 또한, XML은 SGML의 실용적인 기능만을 모은 부분집합 (subset)이라 할 수 있으며, 인터넷상에서뿐만 아니라 전자 출판, 의학, 경영, 법률, 판매 자동화, 디지털도서관, 전자상거래 등 매우 광범위하게 이용될 전망이다.
XML은 월드와이드웹, 인트라넷 등에서 데이터와 포맷 두 가지 모두를 공유하려고 할 때 유용한 방법이라 할 수 있다.
XML은 현재 W3C로부터 웹을 좀더 다양한 목적으로 이용할 수 있도록 하기 위한 도구로서 공식 추천되고 있다

◉웹서비스(web service)

웹서비스는 자기 기술적이며 퍼블리싱이 가능하고, 웹이나 개방 인터넷 표

을 기반으로하는 로컬 네트워크라면 어떤 곳에도 위치할 수 있으며, 호출도 가능한 모듈 응용프로그램이다. 웹 서비스는 HTTP나 SMTP와 같이 흔히 볼 수 있는 인터넷 프로토콜을 통해서 접근할 수 있으며 XML 기반으로 되어 있다. XML 웹 서비스를 사용하는 측은 어떠한 언어나 컴포넌트 모델로 구현되어도 무방하며, 어떠한 운영체제에서 호스팅되이도 무방하다.
사용자 응용프로그램에서 XML웹서비스를 사용하려면 이 웹 서비스를 발견(discovery)해야 한다. 그 다음 XML 웹 서비스가 어떤 것인지 설명(description)이 필요할 것이다. 이 설명은 웹 서비스의 의미를 나타낼 수 있고 XML 웹 서비스를 설명하는 메타데이터로 생각할 수 있다. 마지막으로 사용자는 웹 서비스에 필요한 입력요소를 넘긴 다음, 적절한 결과 데이터를 반환할 수 있게 웹 서비스를 호출(invoke)하게 된다.

●SOAP(Simple Object Access Protocol)

웹서비스가 기술로서 위력을 발휘하기 위해서는 서비스의 발견(UDDI-Universal Description Discovery and Integration) 및 서비스의 기능 결정(WSDL : Web Service Definition Language)과 관련해서 정제된 접근법이 있어야 한다. SOAP이 웹 서비스의 메시징 프로토콜로서 최상의 선택이기는 하지만, 사용할 수 있는 다른 프로토콜이 존재한다. 그러나 일반적으로 웹서비스에서는 메시징 프로토콜로 SOAP이 사용된다. 벤더와 여러 개발자들에 의해 SOAP는 광범위하게 받아들여지고 있다. SOAP스펙은 메시지 교환에 대한 모델을 정의하고 있다. 이 모델에서 '메시지는 XML문서이다', '메시지는 발신자(sender)에서 수신자(receiver)로 이동한다', '수신자는 일련의 체인으로 연결된 형태가 될 수 있다'의 세 가지 기본 개념으로 이루어져 있다. 이 세 가지 개념만 준수하면 SOAP를 충족하는 시스템을 만들 수 있다.

SOAP는 DCOM, CORBA/IIOP, RMI와 같은 기본의 프로토콜을 완전히 대체하지는 못한다. SOAP는 고전적인 메시징과 분산 객체 시스템에 관한 몇몇 기능을 스펙에 넣지 않았다. 예를 들면 활성화(activation), 참조에 의한 객체(object by-reference), 분산 가비지 콜렉션, 그리고 메시지 일괄처리(batching of messages)와 같은 객체 형식의 기능은 일부러 SOAP클라이언트와 서버가 구현되는 인프라에 남겨두었다. SOAP는 객체 모델을 가지고 있지 않으므로 다양한 객체 모델과 언어를 사용해서 구현된 응용 프로그램에 바인딩된다.

◉WSDL(Web Services Description Language)

인터넷상에 있는 컴포넌트의 수백만 조각들이 어떠한 개발 언어상에서도 사용가능하다는 웹서비스를 실현시키기 위해서는 웹서비스는 자기 서술적이어야 한다. 자바 클래스와 COM객체가 자기 서술적인 것처럼 웹서비스도 WSDL을 사용하여 인터페이스와 바인딩을 기술한다. 만약 웹서비스가 그들의 인터페이스를 WSDS을 사용하여 배포하면, 어떠한 웹 클라이언트도 구현상의 세부내역을 알 필요없이, 어떠한 플랫폼이나 운영체제에서도 프로그램적으로 호출할 수 있다. WSDL은 XML, XSD, SOAP, MIME와 같은 개방 표준에서 작성되었으며 플랫폼, 구현, 장치 독립을 현실로 이루어준다.

◉UDDI(Universal Description Discovery and Integration)

웹기반 레지스트리에 대한 표준을 정의한 것이다. 회사들은 이 레지스트리에 자사의 정보, 제공하는 서비스와 서비스를 사용하는 방법에 관한 기술 정보를 등록할 수 있다. 다른 회사나 서비스 제공자의 레지스트리 또한 검색할 수 있다. 이 레지스트리는 폐쇄적일 수 있고, 전역적인 UDDI 표준 레지스트리가 될 수도 있다.

◉멀티미디어 데이터베이스

멀티미디어 데이터베이스는 사용하는 사람에 따라 다른 의미를 가진다. 그에 따라 다양한 정의를 가진다. cd-rom, 단순 멀티미디어 시스템, 주시 요구 비디오, 문서 관리 및 이미지 시스템, 대량의 이진 객체, 객체 지향 데이터베이스 등등이 있다. 멀티미디어 데이터베이스는 멀티미디어를 위한 데이터 타입을 지원하고 조작하는 능력이 있어야 한다. 데이터의 특성상 대용량을 지원해야 하고 효과적으로 저장할 수 있어야한다. 그리고 멀티미디어에 대한 정보 검색 능력이 있어야 한다. 다양한 멀티미디어 데이터를 관리하기 위한 관리 시스템의 역할은 다음과 같다. 데이터의 공간적인 관계성을 위한 공간 데이터 타입과 질의를 지원해야 한다. 사용자가 원하는 데이터를 가져오기 위해 대화식 질의, 피드백을 할 수 있어야 한다. 멀티미디어 데이터의 특징을 자동 추출하고 인덱싱할 수 있어야 한다. 질의를 처리하는 시간을 단축시키기 위해 전통적인 데이터베이스에서 사용되는 B-트리, 해쉬 트리 기반의 인덱싱보다는 공간적, 다차원적인 인덱스를 지원해야 한다.

◉멀티미디어 데이터 인덱싱 방법

html text, xml data, image, video와 같은 반구조화된 데이터(less-structured data)는 WWW같은 인터넷 기술의 확산으로 PC저장공간에 축적할 수 있게 되었다. 반구조화된 데이터는 총괄적으로 멀티미디어 문서(multimedia ducument)라고 부른다. 멀티미디어 문서를 관리하는 데이터베이스는 키워드 기반(keyword-based), 내용기반(content-based)의 검색기능을 지원할 수 있어야 한다. 그러나 관계형 데이터베이스에서 DBMS는 멀티미디어 문서를 BLOB의 스트림 형태의 자료형으로 저장, 관리한다. DBMS는 BLOB

로 저장된 문서가 어떤 내용을 가지고 있는지 해석할 수 없다. 데이터베이스에 저장된 멀티미디어 문서를 분석하고,관리하기 위해서는 다큐먼트 웨이하우스가 필요하다. 문서를 관리하는 웨어하우스가 필요로하는 기능은 다음과 같다. 문서의 내용을 분석하고, 키워드를 추출하고, 동영상, 비디오에서 장면을 추출하는 기능을 제공해야 한다. 추가적인 검색을 위해 문서의 키워드와 속성을 이용하는 쿼리 능력을 제공해야 한다. 사용자가 찾고자하는 문서에 대한 표현을 이용한 내용기반 검색기능을 제공해야 한다. 관련된 멀티미디어 문서를 찾기 위해 키워드기반, 내용기반 검색능력을 조합하여 쿼리를 재정의할 수 있어야 한다. 사용자가 제공한 관점에 따라 멀티미디어 문서를 자동적으로 분류할 수 있어야 한다.

멀티미디어 인덱싱 방법은 데이터에 대한 부수적인 정보에 대한 인덱싱과 데이터의 내용(의 특성)을 표현하는 인덱싱으로 나눌 수 있다. 영화 제목, 스태프, 캐스팅 등은 부수 정보를 볼 수 있고, 줄거리, 씬 구조등은 영화에 대한 내용으로 볼 수 있다.

2.2 기존 연구

멀티미디어 데이터는 복잡하고 다양한 내용을 가지고 있으므로, 그 내용이나, 내용에 대한 특성으로 인덱싱하기가 쉽지 않다. 원격 데이터베이스들에 저장되어 있는 멀티미디어 데이터에 대한 내용을 이용한 통합된 인덱싱 방법은 복잡한 방법과 비용을 필요로 하고 구현하기가 어렵다.

Kansei Retrieval System[1]은 이 문제를 해결하기 위해서, 사용자의 관심을 이용한 인덱싱 방법을 제안했다. 이 시스템의 메타서치 엔진은 데이터의 내용에 대한 인덱싱은 하지 않고, 사용자가 접근한 히스토리를 이용해 데이터의 특성을 추론하고, 이 특성을 이용해 반자동으로 인덱스를 생성한다. 예를 들

이, 영화의 경우 관리자가 특정 장르로 분류하여 검색될 수 있도록 할 수 있다. 그 영화를 관리자가 분류한 장르로 검색할 수 있지만, 사용자가 다른 장르에도 포함된다고 판단할 수 있다. 이렇게 사용자가 결정한 장르도 검색에 이용되도록 하여, 다음 사용자가 원하는 데이터를 검색할 수 있도록 도움을 주게 된다. 그림 1은 kansei retrieval system에 대한 전체적인 흐름을 보여주는 그림이다.

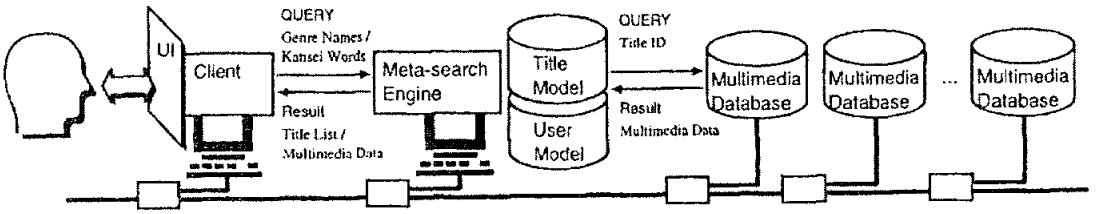

그림 1. Kansei Retrieval System

Kansei word, phrase로 데이터의 내용에 대한 이용자의 생각을 나타내었다. 데이터제공자가 제공한 검색정보를 title model로 하고 이용자의 주관적인 정보를 user model로 하였다. 시뮬레이션에서는 임의로 title model과 user model을 만들어 두 모델사이의 유사도를 이용해 리스트에서 선택을 하고, 재생산정도를 측정하였다.

리스트에서 선택하는 정도를 판단하는 유사도는 다음의 식(1)과 같다.[1]

$$\cos^{p}\theta = \left\{ \frac{\vec{PT_m} \cdot \vec{PU_n}}{|\vec{PT_m}||\vec{PU_n}|} \right\}^{p} \quad \cdots\cdots \text{식(1)}$$

2.3 기존 연구에 대한 문제 제기와 시스템 제안

kansei retrieval system은 멀티미디어 데이터 중 video 데이터를 위한 검색

시스템을 구축하였다. 멀티미디어 데이터를 검색하는데 이용되는 인덱스가 video데이터에 대한 특정 장르로 한정하였기 때문에 사용자의 데이터에 주관적인 느낌을 제한하여 예측하였으므로, 데이터의 특성이 정확하게 유지되지 못하였다. 검색 시스템은 실제 데이터에 대한 분석을 충분히 하지 못하고, 또 데이터간의 상호 관계에 대한 검색 정보가 부족하였다. video 데이터를 대상으로 한 검색 서비스를 할 경우 데이터간의 상호 관계를 표현하지 못했다. 또 성능측정에서 재생산정도가 시간이 지남에 따라 떨어졌다.

본 논문은 멀티미디어 데이터를 포함하거나 가리키고 있는 XML 문서를 위한 검색 시스템을 구축한다. XML 문서를 보관하고 있는 데이터 제공자가 제시한 정보와 사용자가 이용한 검색 질의어를 검색 시스템에서 함께 유지하여 효과적으로 XML 문서를 검색하는 시스템을 제안한다. 실제로 존재하는 XML 문서는 관계형 또는 XML 문서 형식을 처리할 능력이 있는 원격 데이터베이스에 저장되어 있다고 가정한다. 검색 시스템은 원격 사이트에 보관되어 있는 XML 문서에서 가장 많이 이용되는 Element를 저장·관리할 수 있도록 한다. 즉, 원하는 문서를 검색하는데 이용되는 검색어는 XML 문서내에 존재하는 Element들 중 그 문서를 대표할 수 있는 것으로 경로를 포함한다. 이러한 검색어에 사용자의 선호도를 부여하여 검색한다. XML 문서의 특정 Element에 대한 path와 text를 이용해 검색을 할 경우에, 사용자의 선호도를 이용하여 효과적인 문서 검색을 하고 문서간의 상호관계를 예측한다. XML 문서에서 검색에 이용될 Element와 그룹들을 검색 인덱스에서 유지하여 사용자가 입력한 검색어들을 검색 인덱스에 맞게 변환하여 원하는 XML 문서를 검색할 수 있게 한다. 또한, 사용자가 검색에 이용한 Element와 text를 검색 인덱스에 삽입할 수 있게 한다. 성능 측정에서 재생산 정도(reproduction degree)가 일정 수치로 유지될 수 있도록 개선한다.

3. 효율적인 XML문서 검색 시스템의 제안

3.1 XML 문서 검색 시스템

멀티미디어 데이터가 저장된 원격 데이터베이스가 관계형 또는 그 외의 형태로 저장하여 XML 문서를 다루는 경우의 데이터 검색 모델을 제안한다. 원하는 문서를 찾아주기 위해서 사용자가 입력하는 검색어에 대한 선호도를 검색 인덱스에 포함하여 관리하고, 문서간 상호 관계를 연결하고자 한다.

XML 문서에 대한 원래의 정보와 그에 대한 사용자의 주관적인 느낌을 이용하여 다음에 검색될 문서들의 리스트를 예측할 수 있다. 검색 시스템에서는 원격 데이터베이스에 저장되어 있는 XML 문서 내에 존재하는 Element중에서 검색에 자주 이용된다고 판단되는 것들을 보관하고 있다. 사용자는 검색시스템에서 유지되고 있는 Element를 이용하여 리스트를 통해 원하는 문서를 검색할 수 있고, 검색 시스템의 인덱스가 불충분하다고 판단되면 직접 원격 데이터베이스에 저장된 XML 문서를 검색하게 된다. 이 때, 이용한 검색어들은 특정 문서에 대한 사용자의 선호도를 나타낸 것으로 검색 시스템에서 다음 사용자가 이용할 수 있도록 반영하게 된다. 사용자가 원하는 문서를 얻기 위해 많이 사용하는 검색 인덱스일수록 가중치가 높아져서, 다음에 검색되는 후보 리스트에서 상위에 위치할 수 있다.

XML 문서에 대한 사용자의 느낌을 표현하기 위해서는 문서에 대한 기본 정보가 검색시스템에서 유지되어야 한다. XML 문서에 대한 기본정보는 그 문서가 검색되기 위해서 데이터제공자가 등록한 Element와 text이다. 사용자

가 원하는 문서를 검색하기 위해 입력한 검색어는 먼저 검색 시스템을 이용한다. 검색시스템에서는 사용자의 검색어를 element와 text값으로 변환하여 매치되는 관련 문서의 리스트가 작성되면 사용자에게 보여 준다. 만약 관련 문서의 리스트가 작성되지 않으면 원격 데이터베이스에 접속하여 직접 찾게 된다. 검색 서비스를 제공한 후 사용자가 입력한 검색어 중 타당하다고 판단된 것은 같은 Element내에 다른 text값으로 두거나 새로운 Element를 생성할 수도 있다. 사용자가 등록한 검색어는 다음 사용자가 검색을 할 때 관련된 문서를 검색하는데 이용한다.

사용자가 등록한 검색 정보는 원하는 문서를 찾거나 관련된 다른 문서를 찾을 수 있도록 도와줄 수 있다. 사용자가 입력한 정보는 관련된 다른 문서에 대한 정보를 가리킬 수 있다. 특정 문서에 대한 원래의 검색 정보를 공유하는 다른 문서를 검색할 수도 있지만, 사용자가 등록한 검색 정보를 이용하면 이전 이용자의 히스토리로 관련성 있는 문서를 검색할 수 있다.

관련 문서에 대한 정보에서, 사용자의 검색 선호는 해당 XML 문서에 대한 주관적인 느낌을 표현한다. 사용자의 주관적인 느낌은 관련성 있는 문서에 대한 정보를 포함할 수 있다. 특정 문서에 대한 검색 정보가 기본적으로 저장되어 있고 여기에 이전 사용자가 입력한 정보는 기존 문서의 정보와는 다른 내용을 가지지만 다른 문서의 검색 정보와 일치하게 된다. 사용자가 특정 문서를 검색하게 되면, XML 문서에 대한 인덱스에는 다른 사용자들의 선호도가 반영된 인덱스가 있다. 이 인덱스로 관련 문서를 찾을 수 있다. 예를 들어, 주인공의 항목에 사용자가 조연 출연자 이름을 등록하게 되면, 다음 사용자가 검색을 하게 될 때 관련문서로 이 출연자가 주인공으로 등록되어 있는 문서를 검색할 수 있다.

검색 시스템은 관련된 문서를 검색하기 위해서 다수의 Element들을 이용한다. 검색 인덱스는 원격 XML문서의 지시자 아래에, 관련된 Element들이 연결되어 있는 트리 형태를 가진다. 즉 하나의 문서를 대표적으로 표현할 수 있는 검색어들을 유지 관리한다. 검색시스템은 이들 문서에 저장된 Element들의 정보를 저장한다. 필요성과 유효성, 유사성 등의 부가 정보로 해당 검색 정보가 검색 시스템에서 가치가 있는지를 판단한다.

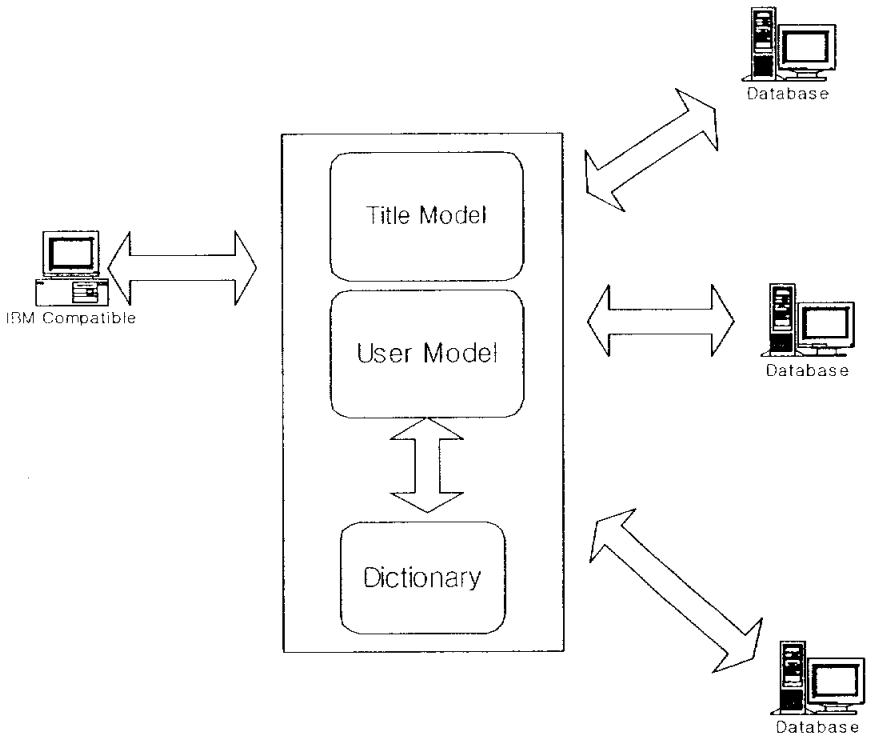

그림 2. 전체 시스템 구성도

3.2 시스템의 구성 요소

3.2.1 Title Model과 User Model

Title Model은 데이터 제공자가 자신의 문서가 효율적으로 검색이 되도록 하기 위해 제시한 검색 정보이다. 원격지에 실제로 존재하는 XML문서의 내용 중 데이터 제공자가 제시하거나 검색시스템에서 선별을 하여 Title Model에 보관한다. Title Model에 보관되는 검색 정보는 내부 사전에 따라 하위 Element로 분류되고 Text 는 원격지의 Text로 한다. User Model은 Title Model을 바탕으로 사용자가 검색에 이용한 검색어들을 등록한 검색 정보이다. Title Model의 Element 분류를 따르고 Text는 사용자가 등록한 값으로 한다. 두 모델은 검색정보를 유지·관리하기 위해 사전을 참조하여 Element를 분류하여 저장한다. User Model에서는 사용자의 기호를 담고 있지만 사용자가 원하는 문서를 검색하기 위한 정보를 정확하게 가지고 있지는 않는다. User Model에서 검색에 이용될 수 있는 것들만을 추출해서 효과적으로 검색할 수 있는 가상의 Title Model를 생성하여 사용자가 원하는 문서를 찾도록 해야 한다.

3.2.2 Element

데이터 제공자가 제시한 정보와 사용자가 입력한 단어들로 관련된 문서를 검색하게 될 때 사용되는 검색 정보이다. XML 문서 내에 존재하는 요소일 수 있고 새로 만들 수도 있다. 원격 데이터베이스에 문서가 등록될 때, 데이터제

공자가 문서내에서 선택하여 검색 정보를 제공하거나, 검색 시스템에서 자동으로 추출하여 검색 시스템에 등록할 수 있다. Element의 분류, 이름, 데이터 값에 대한 정보는 Element 사전을 참조한다.

3.2.3 Element 사전

XML 문서에 대한 데이터 제공자와 사용자의 표현은 매우 다양하다. 데이터 제공자와 사용자의 다양한 표현을 검색 시스템에서 인덱스로 직접 사용하는데는 무리가 있으므로 사전을 두어 입력한 검색어를 검색 인덱스에 해당하는 Element로 변환하여 후보 문서들에 대한 리스트를 생성할 수 있게 한다.

검색 시스템의 Element와 관련된 정보를 관리한다. 검색 시스템에서는 Element를 트리 형식으로 그룹화하고 있다. 검색 시스템에서 사용되는 Element와 값은 사전에 등록된 것으로 한정한다. 이러한 제한은 애매하고 많은 표현으로 인한 부정확한 검색과 과도한 리스트 생성을 방지하기 위해서다. 실제 원격 데이터베이스에 존재하는 문서의 내용을 반영해야 하므로 사전에 등록할 내용은 XML 문서의 Element, Text, 트리 구조를 기반으로 한다. 데이터 제공자나 사용자가 사전에서 변환할 수 없는 검색 정보를 사용할 경우 입력한 검색 정보를 사전에 임시로 등록할 수 있다. Element간의 포함 관계를 나타내는 그룹 정보는 관련된 Element를 계층적으로 연결해주며, Element를 이용한 검색 인덱스의 생성, 수정, 삭제를 하는 경우에 이용된다. 그룹간 중복되는 Element는 서로 공유하여 공간 낭비를 해소한다.[2] 사용자가 등록한 Element는 사용량이 많을 경우 정식으로 등록되지만 그렇지 않으면 자동 삭제한다.

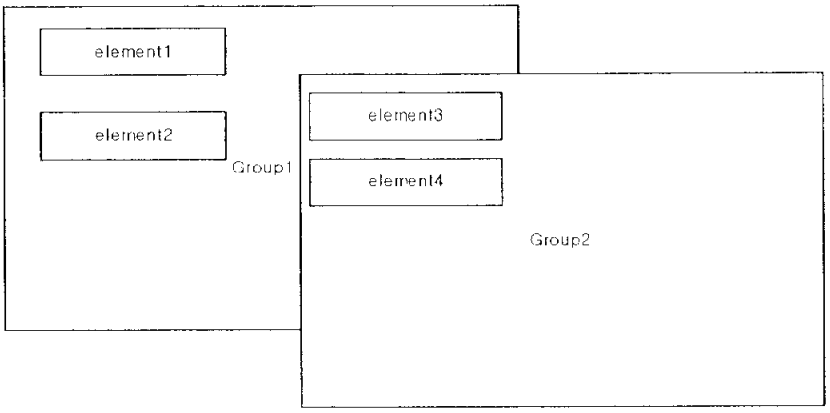

그림 3. 사전의 예

3.3 검색 시스템에서의 인덱스 관리

3.3.1 검색 시스템을 위한 Element 선택

원격 데이터베이스에 XML 문서가 생성되면 검색시스템에서는 해당 문서가 검색되기 위한 Element들이 생성되어야 한다. 데이터 제공자가 검색 정보를 제공하거나 검색시스템의 추천을 받을 수 있다. 데이터 공급자나 검색시스템에 의해 제시된 검색 정보는 사전에 등록되어 있는 Element로 변환하여 지정한다. 추천에 의해 등록을 할 경우 검색 시스템은 XML 문서내의 Element중 검색 시스템에서 사용빈도가 높은 Element를 선택한다. 검색 시스템은 관련된 문서가 검색될 수 있도록 필요한 Element만 관리할 수 있어야 한다. 관련이 없거나 중복된 Element는 문서 검색에 도움이 되지 않으므로 검색시스템에서 존재하지 않도록 해야 한다.

검색 시스템에서 Element의 사용빈도가 적은 Element는 제거를 한다. 불필요한 검색 정보가 과도하게 색새되는 것을 방지하고, 사용자가 검색하고자 하

는 문서 리스트가 효과적으로 만들기 위해서는 사용 빈도가 적거나 부정확한 검색 정보는 삭제한다. 예를 들어 Element의 속성(attribute)에서 유효성이 기준 이하이면 제거를 한다.

3.3.2 검색 정보에 대한 그룹핑

검색 시스템은 유효한 XML 문서를 검색하기 위해서 이용되는 Element를 그룹으로 묶어서 관리한다.

데이터 제공자가 제시한 검색 정보를 사전을 이용해서 등록되어 있는 Element 이름과 속해있는 그룹으로 검색 시스템에 등록하도록 한다. 문서에 대한 지시자를 root Element에 대응시키고 그 하위에 데이터 제공자가 제시한 검색 정보를 바탕으로 사전에서 해당 Element를 링크시켜 문서에 대한 인덱스를 생성한다.

3.3.3 검색 시스템에서의 Element 구성

검색 시스템에서 문서에 대한 검색정보를 관리하기 위해서 XML 문서형식인 트리 구조에 대응하는 검색 정보를 구성할 수 있게 한다. 사전을 참조하여 같은 성질을 가지는 Element를 하나의 그룹으로 묶고, 하위의 성질에 속하는 Element를 하위 그룹으로 연결시키는 분류를 한다. 검색 인덱스로 관리되는 Element중에는 사용자에 의해 자주 이용되는 것도 있고 그렇지 않은 것도 있다. 자주 이용되는 검색 정보는 다음에 문서를 검색할 때 먼저 검색할 수 있도록 하고 그렇지 않은 검색 정보는 삭제를 한다. 검색정보에 대한 부가적인 것은 속성으로 따로 관리한다. 사전을 이용한 검색 인덱스 구성은 그림 4와 같다

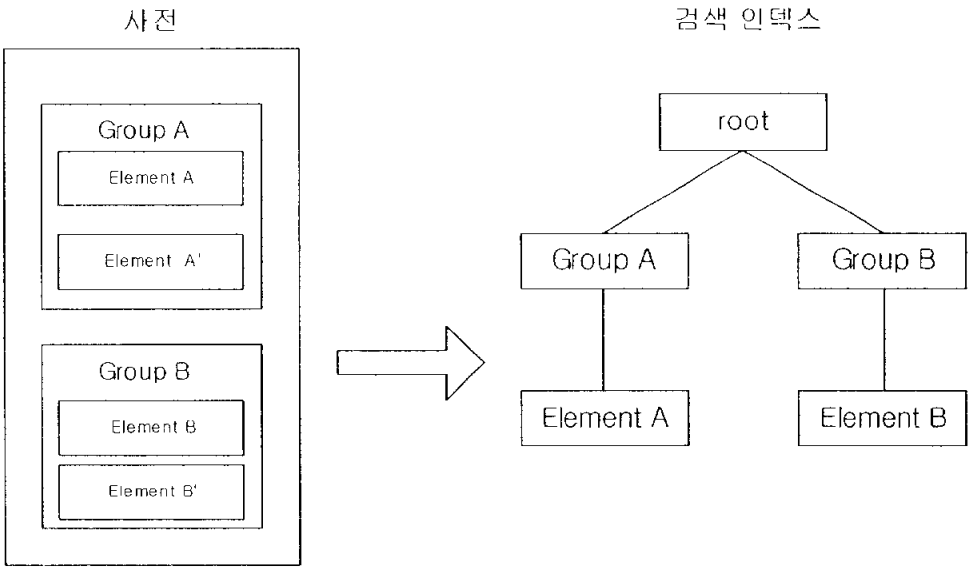

그림 4. 사전을 이용한 검색 인덱스 구성

3.3.4 Element에 대한 속성

Element에 대한 속성은 문서 검색에 이용되는 Element의 부차적인 정보를 저장한다. 검색을 할 때의 검색 빈도에 따른 리스트 내 문서 지시자의 순서, 데이터 제공자와 사용자가 제시한 Element의 구별, Element의 실효성, 다른 Element와의 관련성 등을 판단할 때 이용된다. 대표적인 속성은 필요성, 유효성, 유사성이 있다. 속성에 대한 값은 정규화한 값으로 크기는 (0, 1)로 한다.

필요성은 데이터제공자나 시스템 관리자가 직접 입력한 경우 값을 1로 설정한다. 문서가 검색되기 위해 반드시 필요한 요소라고 판단하여 입력한 것이므로 검색시스템에서 계속 유지될 수 있도록 한다. 사용자가 등록한 element인 경우 검색에 필요한 요소인지 확인할 수 없으므로 0과 1사이에서 미리 설정해

는 값으로 한다. 사용자가 등록한 Element가 계속 검색에 이용이 되고 있다면, 필요성 정도를 계속 높여 준다. 필요성 정도가 감소하여 0이 되면 검색이 이용이 되지 않는 것이므로 검색시스템에서 삭제하도록 한다.

유효성은 관련된 XML 문서의 리스트를 생성하여 정렬을 할 때 필요하다. 사용자에게 관련 문서 지시자의 리스트를 제공하려고 할 때 사용자의 입력을 이용해서 Element들을 찾은 다음 유효성 정도를 우선 순위로 하여 문서들을 정렬하여 리스트를 만든다. 생성된 리스트에서 사용자가 관련 문서를 선택하게 되면 선택된 문서에 대한 element의 유효성 정도를 증가시키고 선택되지 않은 Element의 유효성 정도는 감소시킨다. 생성된 리스트에서 사용자가 많이 이용할수록 유효성이 증가하게 된다.

유사성은 원격 xml 데이터베이스에 저장된 문서의 Element는 아니지만 사용자가 검색에 이용한 입력 어구와 원래 값과의 관련성을 나타낸다. 사용자가 특정 문서를 검색하기 위해 입력한 검색어는 원래 문서의 내용과 일치할 수도 있지만 원래의 내용과는 다르게 판단할 수 있다. 사용자가 등록한 검색 정보와 원래 문서와의 관련성을 유사성으로 판단한다. 사용자가 특정 문서를 찾고자 할 때 관련성이 클수록 1에 가까운 값을 가진다.

Element와 관련된 텍스트와 속성의 예는 그림 5와 같다.

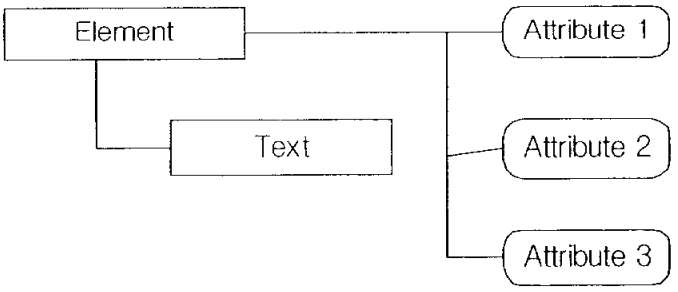

그림 5. Element와 관련된 텍스트와 속성

3.4 사용자의 검색어와 검색 인덱스 사이의 매칭 방법

검색시스템이 사용자의 입력을 받아들여 관련된 문서의 리스트를 만들기 위해서는 여러 가지의 매칭 방법을 이용할 수 있도록 한다.

3.4.1 확대 매칭

검색 범위로 리스트를 생성하게 될 때, 그 범위에 근접해 있는 Element도 검색 리스트에 등록할 수 있게 한다.

(예) 등장 인물이 많다

> 70인 이상으로 범위를 정한 경우에, 68, 69명이 등장하는 멀티미디어도 검색 리스트에 등록

3.4.2 우선 순위 매칭

각 element의 유효성, 유사성 등의 정도를 이용하여 리스트를 정렬한다. 이전의 사용자에 의해서 가장 많이 이용된 검색정보를 최상위에 올려 다음의 사용자가 문서를 검색하는데 도움을 주게 한다.

3.4.3 선호도에 의한 매칭

사용자가 입력한 검색어에 우선순위를 부여하여 리스트를 정렬한다. 현재의 사용자가 여러 조건을 동시에 제시하면, 검색어의 중요도에 따라 리스트를 생성하도록 한다.

3.4.4 조건 매칭

선 조건을 만족한 리스트에서 다음 조건을 만족하는 리스트를 생성한다. 처음의 검색어로 리스트를 만든 다음에 다시 검색어를 입력하여 이전의 리스트의 크기를 줄여 리스트를 다시 만든다.

3.4.5 검색 시스템에서의 Element 트리 확대 및 축소 과정

사용자는 관련된 XML 문서의 리스트를 검색하기 위해, 하나 이상의 검색어(이름, 값)를 입력한다. 사용자가 입력한 검색어들을 검색 시스템에 맞는 Element의 이름과 값으로 변환하여 매칭되는 XML 문서의 리스트를 찾는다. 관련 XML 문서의 리스트를 검색하는 과정에서 검색어가 변환된 Element의 확대나 축소과정이 필요하다.[3] 사용자는 비전문가의 입장으로 검색을 하게 되므로 검색시스템에 대한 이해없이 검색어를 입력한다. 이렇게 입력한 검색어를 검색시스템에 있는 Element와 대응하기 위해서는, 검색 시스템에서 그룹지어져서 관리되는 Element와 같이 사전을 참조하여 그룹핑하는 과정이 필요하다. 매칭과정에서 검색어로 찾은 Element의 분류 깊이 보다 작은 경우는 확대를 하여 매칭이 되는지 확인을 해야 한다. 그림 6은 Element 트리의 확대 및 축소의 예를 나타내고 있다.

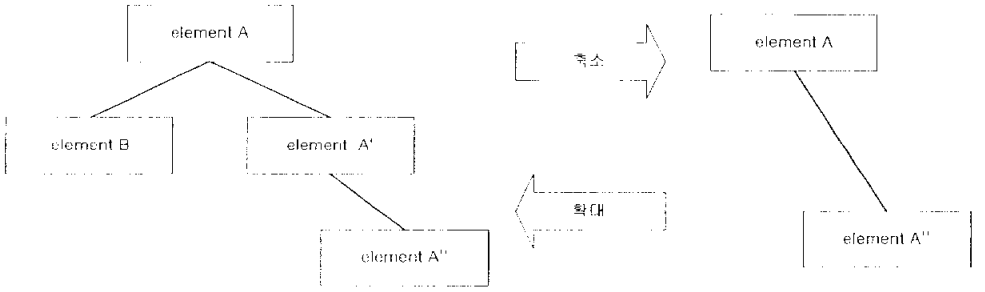

그림 6. Element 트리의 확대 및 축소

3.4.6 검색어의 변환

사용자가 입력한 검색어를 검색시스템의 Element에 맞게 변환하는 과정이 필요하다.[4] 사용자는 자신의 느낌으로 검색어를 입력하므로 직접 검색하기에는 충분하지 않다. 사용자가 입력한 검색어는 검색 사전을 이용해서 Element의 이름과 값, 속하는 그룹으로 변환시킨다. 사용자가 입력한 검색어와 검색 사전사이의 매치는 따로 구현할 수 있도록 한다.

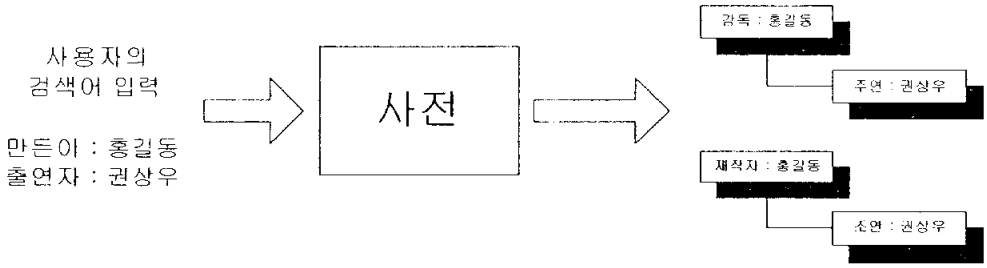

그림 7. 사용자가 입력한 검색어의 변환

3.4.7 잠재적인 Title Model 생성

사용자가 입력한 검색어를 이용해 검색을 하게 될 때 사용되는 검색 모델이다. 미리 만들어진 Title Model은 데이터 제공자와 관리자에 의해 만들어져 수정을 거의 하지 않는 고정된 모델이다. 이 Title Model에 사용자의 검색 선호를 포함하고 있는 User Model을 결합한 검색 모델로 이전의 사용자의 선호도가 반영된 검색을 할 수 있게 한다. User Model내의 검색 정보들은 사용자가 입력한 검색어가 등록되어 있어 실제 문서 검색에 불필요한 것도 있다. 이 User Model에 존재하는 정보들 중에서 선별하여 실제 검색에 활용될 수 있는 잠재적인 Title Model 생성 작업이 필요하다. 잠재적인 Title Model은 기본적으로 Title Model로부터 생산된다. 이 잠재적인 Title Model과 데이터 제공자가 제시한 Title Model로부터 재생산도가 계산된다. 잠재적인 Title Model을 만들어낼 때는 Title Model과 근접한 모델이 되도록 한다. 예를 들어 User Model에서 Title Model에 비해 선호도가 너무 높은 검색 인덱스를 선택할 경우 재생산도가 낮게 측정된다. User Model에서 Title Model의 검색 인덱스와 유사한 벡터 크기를 가지는 검색 인덱스를 선택하게 되면 재생산도는 정규화된 값에 근접하게 된다.

3.5 최종 XML 문서 선택

검색된 리스트에서 사용자가 최종적으로 선택을 하게 되면, 문서를 저장하고 있는 원격 데이터베이스에 요청을 하여 데이터를 받을 수 있다. 검색 엔진에서는 원격 데이터베이스의 서비스 방식에 따라 요청 방법을 바꾸어서 요청을 할 수 있어야 한다. 해당 데이터베이스가 XML 문서를 관리할 수 있다면 XPATH를 이용해서 데이터를 가져올 수 있다. 해당 데이터베이스가 관계형

데이터베이스인 경우 sql문을 생성하여 데이터를 요청한다. 사용자가 최종적으로 선택한 문서를 가져오는 과정은 그림 7과 같다.

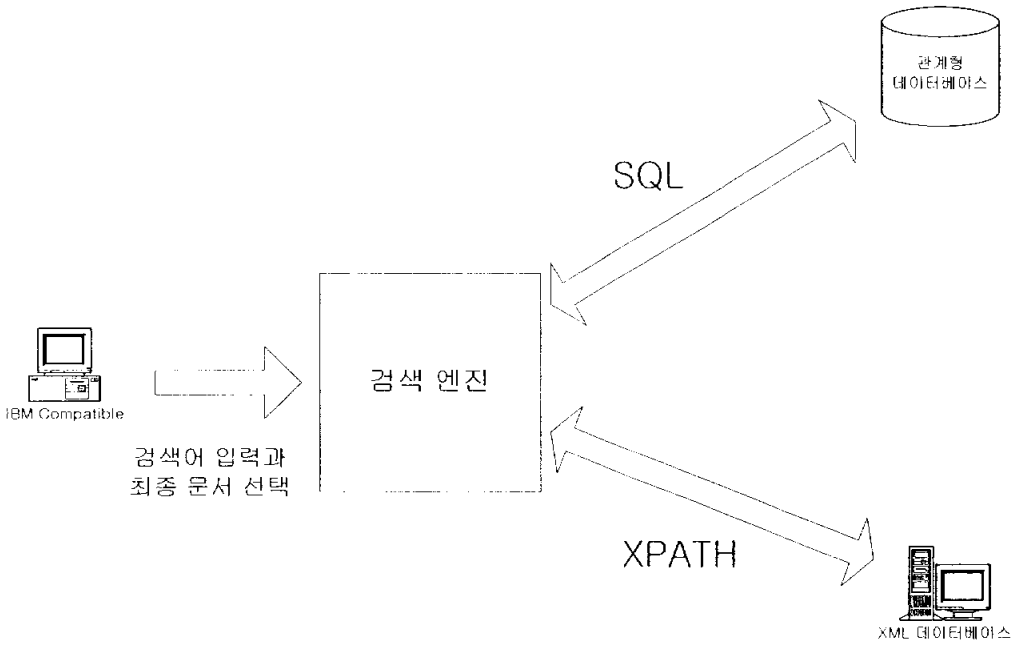

그림 8. 사용자가 최종적으로 선택한 문서를 가져오는 과정

3.6 성능 평가

문서의 실제 검색정보를 유지하는 Title Model과 User Model이 필요하다. Title Model은 서로 다른 데이터 제공자가 제공하는 정보이고, User Model은 사용자들의 관심을 반영한 정보이다. 이 두 모델은 벡터공간에서 수치를 표현할 수 있다. 시간의 구간을 c_i로 할 경우에, c_0인 경우의 Title model은 T_{c_0} = (1,0)이 된다. User model은 사용자의 선택에 따라 크기가 달라지므로 U_{c_i} = (0.9, 0.7)로 표현할 수 있다. 이 두 모델을 이용해서 데이터의 특성과 사용자의 관심사이의 관계를 설명한다.

Title model과 User model로 다음에 이용될 각 model을 예측할 수 있다. 식 2와 식 3과 같다

$$\vec{U_n}(C_n+1)=\frac{C_n\times \vec{U_n}(C_n)+\vec{T_m}(c_m)}{C_n+1} \quad \cdots\cdots 식(2)$$

$$\vec{T_m}(C_m+1)=\frac{C_m\times \vec{T_m}(C_m)+\vec{U_n}(c_n)}{C_m+1} \quad \cdots\cdots 식(3)$$

위의 예측값은 자신의 값을 서로에게 피드백하여 구하게 된다. 여기서 구한 예측값은 각 Model에서의 특성과 관심을 나타내는 잠재적인 값으로 사용할 수 있다.

T_m : 데이터의 실제 내용만으로 데이터의 특성을 반영하는 벡터 (1,0), (0,1)

PT_m의 잠재적 특성의 최대값과 관련이 있으므로 크기는 1로 고정된다.

T_m이 1로 고정이 되어 있어야 PT의 최대값이 1이 될 수 있다.

U_n : 사용자가 등록한 내용의 관심을 반영하는 벡터이다. 시간이 지남에 따라

사용자의 선호도에 따라 그 값이 변화한다

(0.9, 0.3), (0.3, 0.4)

시뮬레이션을 위해 잠재적인 타이틀 모델인 PT(Potential Title Model)를 만들어 낸다. PT는 사용자가 입력한 검색어로 검색하는 실제 검색 인덱스와 같다. 검색 시스템은 사용자가 등록한 검색정보를 다음 사용자에게 반영하기 위해 기존의 Title Model에 User Model중 검색 정보로 가치가 있다고 판단되는

성을 추출하여 잠재적인 Title Model를 생성한다.

PT_m은 T_m과 U_n으로부터 추측하도록 한다. 사용자의 관심도에 따라 잠재적인 특성의 백터를 결정하도록 하였나. U_n은 일정 크기로 조건에 따라 변하도록 하였다.

검색 시스템의 재생산도(reproduction degree)를 계산한다. 재생산 정도를 이용해서 Title model의 T가 잠재적인 특성 PT를 생산해 내는지를 알 수 있다. PT는 Title Model과 User Model의 피드백 작용으로 구할 수 있다. Title Model의 크기 변화는 User Model에서 필요한 검색 정보를 선별하는 기준이 되기 때문이다. 특정 element의 유효성 정도와 이 특정 element와 관련하여 생성된 element의 유효성 정도를 이용하여 모델의 재생산 정도를 시뮬레이션 한다. 재생산정도를 계산하는 공식은 다음과 같다.

$$\cos\theta = \frac{\overrightarrow{T_m(C_m)} \cdot \overrightarrow{PT_m}}{|\overrightarrow{T_m(C_m)}||\overrightarrow{PT_m}|} \quad \cdots\cdots 식(4)$$

데이터 제공자가 제공한 검색 정보는 각 레벨당 50개, 깊이 5로 하고, 사용자가 제시한 검색 정보는 각 element당 3개로 제한하여 재생산 정도를 체크한다. XML 문서를 검색하기 위해 데이터 제공자가 제시한 검색 정보를 Titile Model로 하고 사용자의 입력으로 만들어진 검색 정보를 User model로 한다. 잠재적인 Title Model생성과 재생산도 측정에 대한 예는 그림 9와 같다.

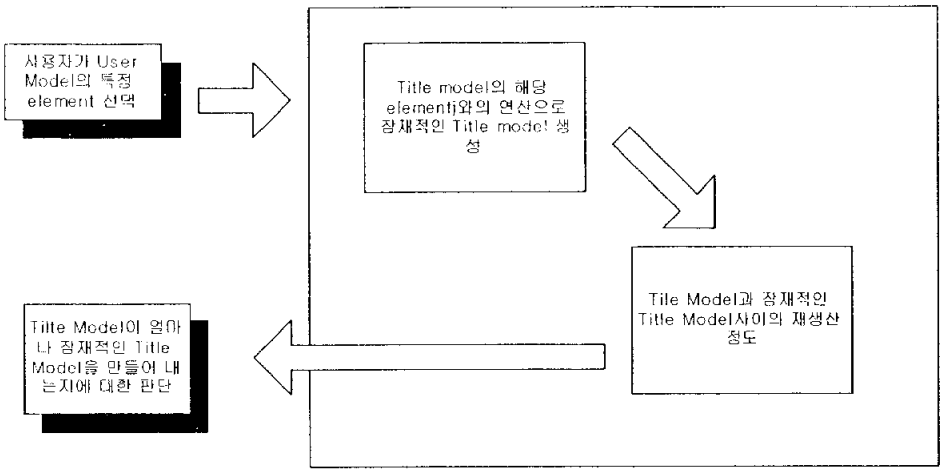

그림 9. 잠재적인 Title Model생성과 재생산도 측정

시뮬레이션에서는 최적의 재생산도를 만들기 위해 User Model로부터 적절한 Element를 추출하여 잠재적인 Title Model을 만든다. 잠재적인 Title Model은 User Model로부터 유효한 검색 빈도를 가지는 Element를 선택하여 검색에 이용되게 하므로 잠재적인 Title Model에 따라 재생산도(Reproduction Degree)가 결정된다. 다음 표 1, 2, 3은 Title Model이 일정크기를 가지는 벡터일 때 잠재적인 Title Model의 크기가 일정하게 변화하는 경우의 재생산도를 측정한 예다.

PT	Reproduction Degree
(0.00, 1.00)	0.707
(1.00, 0.95)	0.999
(0.95, 0.90)	0.999
(0.90, 0.85)	0.999
(0.85, 0.80)	0.999
(0.80, 0.75)	0.999
(0.75, 0.70)	0.999
(0.70, 0.65)	0.999
(0.65, 0.60)	0.999
(0.60, 0.55)	0.999
(0.55, 0.50)	0.998
(0.50, 0.45)	0.998
(0.45, 0.40)	0.998
(0.40, 0.35)	0.997
(0.35, 0.30)	0.997
(0.30, 0.25)	0.995
(0.25, 0.20)	0.993
(0.20, 0.15)	0.989
(0.15, 0.10)	0.980
(0.10, 0.05)	0.948

표 1. Title Model이 (0.5, 0.5)일 때, PT의 변화에 따른 재생산도

PT	Reproduction Degree
(0.00, 1.00)	0.650
(1.00, 0.95)	0.998
(0.95, 0.90)	0.998
(0.90, 0.85)	0.998
(0.85, 0.80)	0.998
(0.80, 0.75)	0.999
(0.75, 0.70)	0.999
(0.70, 0.65)	0.999
(0.65, 0.60)	0.999
(0.60, 0.55)	0.999
(0.55, 0.50)	0.998
(0.50, 0.45)	0.999
(0.45, 0.40)	0.999
(0.40, 0.35)	0.999
(0.35, 0.30)	0.999
(0.30, 0.25)	0.999
(0.25, 0.20)	0.999
(0.20, 0.15)	0.997
(0.15, 0.10)	0.992
(0.10, 0.05)	0.970

표 2. Title Model이 (0.7, 0.6)일 때, PT의 변화에 따른 재생산도

PT	Reproduction Degree
(0.00, 1.00)	0.496
(1.00, 0.95)	0.971
(0.95, 0.90)	0.971
(0.90, 0.85)	0.971
(0.85, 0.80)	0.972
(0.80, 0.75)	0.972
(0.75, 0.70)	0.973
(0.70, 0.65)	0.973
(0.65, 0.60)	0.974
(0.60, 0.55)	0.975
(0.55, 0.50)	0.976
(0.50, 0.45)	0.977
(0.45, 0.40)	0.980
(0.40, 0.35)	0.982
(0.35, 0.30)	0.984
(0.30, 0.25)	0.987
(0.25, 0.20)	0.992
(0.20, 0.15)	0.997
(0.15, 0.10)	0.992
(0.10, 0.05)	0.998

표 3. Title Model이 (0.4, 0.7)일 때, PT의 변화에 따른 재생산도

표 1, 2, 3에서 보듯이 잠재적인 Title Model의 변화폭이 크거나 그 크기가 작은 경우 재생산도가 낮아짐을 보여준다. 또한 높은 재생산도를 구하기 위한 PT의 선택은 Title Model의 크기와 변화폭에 따라 달라진다. 재생산정도를 향상시키기 위해서는 검색시스템에서 잠재적인 Title Model이 급격하게 변화하는 것을 막고 그 크기가 너무 작아지는 것을 방지해야 한다. 또한 Title Model에서의 벡터 크기에 따라 User Model에서 선택할 Element를 선별해서 선택해야 높은 재생산도를 얻을 수 있다.

실험은 리스트에서 무작위선택, 우선순위 선택, 복수 선택의 방법을 이용했다. 선택이 되는 것은 그 Element의 유효성이 증가되고, 나머지는 유효성이

감소되게 하였다. 그리고 유효성이 작은 특정 User Model은 제거하였다. 실험에 이용한 조건은 다음 표 4와 같다.

Number of users	100
Number of data	100
Number of dimenstions	2
Number of repetition	1,000

표 4. 실험 조건

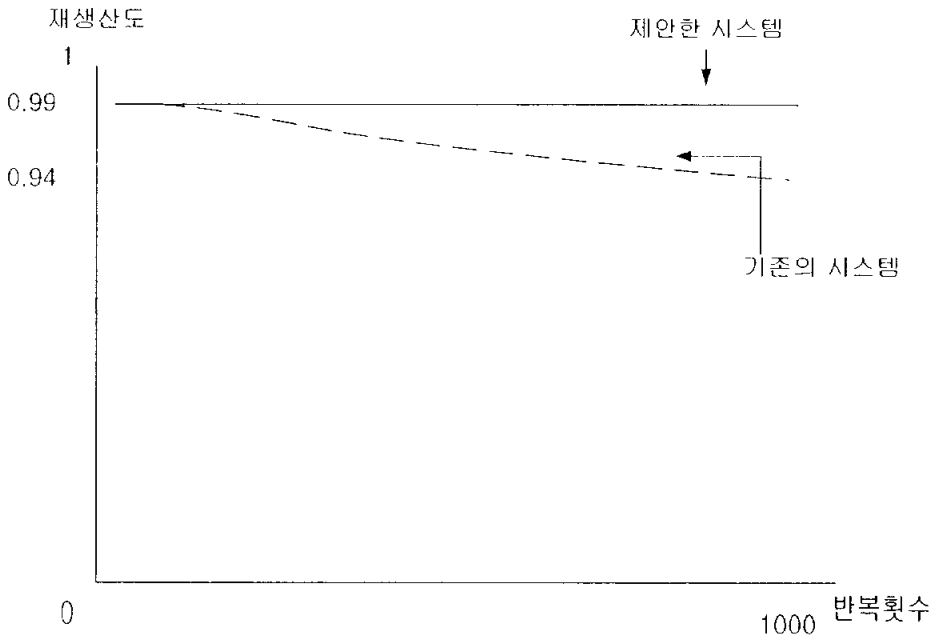

그림 10. reproduction degree의 측정

그림 10에서 반복횟수는 100명의 사용자가 한 번 문서를 선택했을 때의 재생산도를 측정하여 1000번을 각각 측정한 것이다.

테스트에서 User Model에서 무작위로 선택해서 잠재적인 Title Model을 생성할 경우 재생산도가 0.94로 낮게 측정되었다. 즉, User Model에서 낮은 유효성을 가지는 Element를 잠재적인 Title Model에 포함시킬 때는 재생산도가 낮아진다. User Model에서 유효성이 낮은 Element를 제거하고 높은 유효성을 가지는 Element를 잠재적인 Title Model에 포함시킬 경우에는 재생산도가 0.99로 높게 평가되었다.

잠재적인 Title Model의 변화폭과 크기의 조절로 기존 연구에 비해 재생산도를 개선하였고 시간의 변화에 따른 크기의 감소가 없앨 수 있었다.

4. 결론

인터넷이 대중화되고 전송 방법이 발달함에 따라, 이용자들은 다양한 형식의 문서를 인터넷을 통해 접근할 수 있게 되었다. 본 논문은 원격지에 존재하는 멀티미디어 문서중의 하나인 XML 문서를 검색하기 위해 사용자가 등록한 검색정보를 검색 인덱스로 이용하여 다음의 사용자가 활용할 수 있도록 하였다. 사용자가 등록한 검색정보를 User Model로 하여 잠재적인 Title Model을 만들어 내고 실제의 Title Model에 대한 재생산 정도를 측정하였다.

제안한 모델에서 잠재적인 Title Model은 User Model의 변동량에 의해 영향을 받는다. 이 User Model이 효과적으로 사용되면서 변동량이 적어질 수 있도록 하는 연구가 필요하다. 또한 사용자가 입력한 검색어를 검색 인덱스에 대응하는 검색어로 변환하는 연구가 필요하다.

참고문헌

[1] Takashi MITSUISHI, Jun SASAKI, Yutaka FUNYU, A Design of A Kansei Rertrieval System for Distributed Multi-meida Databases, IEEE, 2001

[2] Thomas Kunkelmann, Reberto Brunelli, Advanced Indexing and Retrieval in Present-day Content Management Systems, IEEE, 2002

[3] Didier Dubois, Fuzzy Logic Techniques in Multimedia Database Quering: A Preliminary Investigation of the Potentials, IEEE, 2001

[4] Noriazu Ikoma, Hiroshi Maeda, Yukihiro Kobayashi, Tomomi Tabara, Daisuke Yamamoto, Building Exterior Design System by Hierarchical Combinatin Fuzzy Model, IEEE, 2001

[5] Branko Milosavljevic, Zora Konjovic, Design of an XML-Based Extensible Multimedia Information Retrieval System, IEEE, 2002

[6] Alex G. Buchner, Sarabjot S. Anand, David A. Bell, John G. Hughes, A Framework for Discovering Knowledge from Distributed and Heterogeneous Databases, IEE, 1996

[7] Ian Newman, Applying Experiences of Providing Effective Information Identification and Retrieval for Predominantly Text Based Distributed Databases to Multimedia Databases, IEE, 1998

[8] Hiroshi Ishikawa, Manabu Ohta, Querying Web Distributed Databases for XML-based E-businesses: Requirement Analysis, Design, and Implementation, IEEE, 2001

[9] Patrick Cauldwell, Rajesh Chawla, Vivek Chopra, Professional XML Web Services, Wrox출판사, 2001

[10] Hiroshi Ishikawa, Manabu Ohta, Koki Kato, Document Warehousing: A Document Intensive Application of A Multimedia Database, IEEE, 2001

감사의 글

본 논문이 이루어지기까지 세심한 배려와 지도로 이끌어주시며 학문의 길을 가르쳐주신 조우현 교수님께 한없는 감사를 드립니다. 그리고 본 논문을 심사해 주시고 보완해주신 시경룡 교수님, 신봉기 교수님께 진심으로 감사드립니다. 또한 학위과정동안 많은 가르침으로 주셨던 학과의 여러 교수님들께도 감사드립니다.

대학원 생활을 하는 동안 자유로운 분위기와 편한 마음으로 공부를 하도록 도움을 주신 박희숙 누나에게 감사드립니다. 또한 물심양면으로 도와준 연구실의 김성록, 이상호, 김락기, 김상훈에게도 감사드립니다. 또한 조언을 아끼지 않은 박창수 선배님과 이석주에게도 감사드립니다. 건강에 조언을 준 전희성에게도 감사드립니다.

지금도 고생하시는 어머님께 정말로, 진짜로 감사드립니다.

2004년 6월 정성주 드림