

홈서버 환경에서 메타데이터 기반의 검색 설계

김재원⁰, 김진형, 최오훈
고려대학교 컴퓨터학과
{neoi⁰, koolmania, pens}@software.korea.ac.kr

A design of metadata based search method in Homeserver

Jae-Won Kim⁰, Jinhyung Kim, O-Hoon Choi
Dept. of Computer Science & Engineering, Korea University

요 약

유비쿼터스 환경의 시발점인 홈 네트워크 기술은 홈서버를 중심의 맥내 모든 가전기기를 하나의 네트워크로 연결함으로써 사용자가 언제 어디서 단말 기기를 이용하여 맥내의 상황을 모니터링하고 모든 기기를 제어할 수 있게 해주는 기술이다. 홈 네트워크의 핵심인 홈서버는 다양한 가전기기로부터 다량의 멀티미디어 콘텐츠를 저장하며, 다른 홈서버 또는 모바일 단말을 이용하여 검색과 공유가 가능하다. 그러나, 홈서버와 각 모바일 단말이 가지는 콘텐츠에 대한 메타데이터는 일관성을 가지지 않기 때문에 콘텐츠 검색이 용이하지 않고, 사용자가 직접 메타데이터 간의 유사성을 파악해야하는 문제점이 있다. 또한 메타데이터 간의 이질성으로 인해 이종 기기들 간의 콘텐츠의 검색, 공유 또한 힘들다. 따라서 홈서버들 간의 일관적인 데이터 입력과 공유를 위하여 멀티미디어 콘텐츠에 일반적으로 적용되는 Mpeg-7 기반의 메타데이터 레지스트리 설계를 제시한다. 또한 이질적인 메타데이터를 지니는 홈서버-모바일 단말 간의 콘텐츠 공유를 위하여 워드넷 기반의 온톨로지 설계를 본 논문에서 제시한다.

1. 서 론

인터넷 사용의 급속한 확대와 관련 기술의 발전으로 인터넷은 가장 보편화된 정보 전달의 매체가 되었고, 홈네트워크의 구축 [1][2][3]으로 인터넷 사용 장소는 회사 또는 학교에서 밖으로 이동하고 있다. 특히 디지털 카메라, 캠코더와 이를 저장하는 DVD를 통한 다량의 멀티미디어 콘텐츠의 확산과 이를 여러 다른 맥과 공유하려는 새로운 멀티미디어 서비스 패러다임이 제시되고 있다.

다량의 멀티미디어 콘텐츠는 홈서버, PDA와 핸드폰과 같은 모바일 단말을 통해 검색과 변경이 가능하다. 그러나 홈서버와 각 모바일 단말이 가지는 콘텐츠 데이터 속성은 일관성을 가지지 않아 콘텐츠 검색이 용이하지 않고, 사용자가 직접 내용을 처리하는 부담이 많이 발생하고, 콘텐츠를 공유하기 힘들다. 따라서 이 논문에서는 메타데이터[6]를 이용한 홈서버와 모바일 단말 간의 콘텐츠 검색 및 변경을 위한 설계와 콘텐츠 공유를 위해 메타데이터를 기반으로 하는 온톨로지 설계를 제시한다. 또한 멀티미디어 콘텐츠의 메타데이터를 활용하기 위해 현재 멀티미디어 콘텐츠의 메타데이터 표준으로 흔히 이용되는 MPEG-7(ISO/IEC 15938)[7]의 메타데이터를 사용한다.

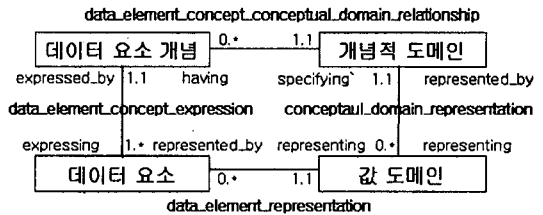
2. 관련 연구

2.1 메타데이터 레지스트리

정보 공유를 위한 데이터 표준화 작업의 일환으로 ISO/IEC 11179 표준[6]은 데이터의 의미, 구분, 표현을 표준화하는 프레임워크를 메타데이터 레지스트리로 제시한다. 메타데이터 레지스트리 데이터 모델은 데이터 요소 개념 (data element concept), 개념적 도메인 (Conceptual domain) 과 값 도메인 (value domain), 데이터 요소(data element)로 구성된다. 데이터 요소는 객체부류 (Object class), 특성 (Property), 표현 (Representation) 의 세 가지 구성요소로 이루어진다. 데이터 요소 개념 (data element concept)은 개념적 도메인, 값 도메인 과 데이터 요소의 3가지 구성요소와 결합하여 특정한 의미적 개념을 가지면서 좀 더 구체화 될 수 있다.[그림 1]

위 메타데이터 레지스트리의 데이터 모델 개념에 따라 콘텐츠도 특정한 데이터 요소 개념으로 재정의[5]가 가능하며, 데

이터 레지스트리에 저장된다. 이 논문에서는 이렇게 정의된 메타데이터를 워드넷 형식으로 홈서버 내에 온톨로지를 구성한다. 이와 같이 홈서버 내에 온톨로지를 구성함으로써 홈서버와 모바일 단말간, 홈서버와 PC간의 콘텐츠 검색 및 변경을 가능하게 한다.



[그림 1] ISO/IEC 11179 메타데이터 레지스트리를 위한 데이터 모델

2.2 MPEG-7 메타데이터

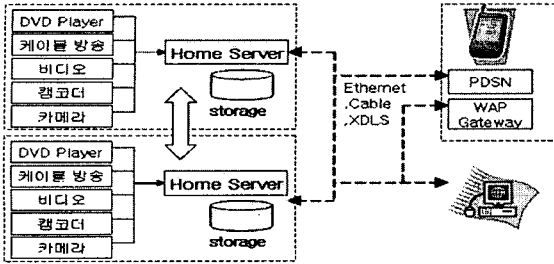
기존의 MPEG-1/2/4 등은 오디오, 영상 데이터의 압축이 목적이었지만, MPEG-7 (ISO/IEC 15938)[7]은 데이터 그 자체가 아닌 데이터의 내용에 대한 표현방법인 '메타데이터'를 다루며, 오디오, 영상 등의 멀티미디어 콘텐츠를 위한 XML 스키마를 이용한 키워드 기반의 표현 및 검색을 효율적으로 지원하는 표준이다.

메타데이터 레지스트리의 데이터 모델 개념을 이용하여 MPEG-7에서 사용될 데이터 요소들은 메타데이터 레지스트리에 저장된다. 일반적으로 사용되는 데이터 요소들은 Title, Creator, Subject, Date, Format 등[표1]이 있다. 그러나 콘텐츠의 데이터 요소들이 정의되었음에도 불구하고, 많은 사용자에 의해서 생성된 다량의 콘텐츠는 같은 의미의 데이터 요소를 가지는 데이터임에도 불구하고, 홈서버-모바일 단말 간의 검색 및 변경이 불가능하다. 예를 들어 [그림 2]의 홈서버와 모바일 단말의 연결 구성에서 MPEG-7에서는 멀티미디어 제목을 'Title'이라고 정의하고 있지만 이를 준수하지 않는 다른 홈서버나 단말에서는 '이름'이라고 정의될 수도 있다. 이 경우 'Title'과 '이름'은 같은 의미를 지니는 메타데이터 요소임에도 불구하고 시스템 측면에서는 이들을 서로 다른 메타데이터 요소로

* 본 논문은 2005년도 한국전자통신연구원, FTTH 기반 홈서버용 양방향 개인용 고품질 콘텐츠 브로드캐스팅 시스템 설계 프로젝트 지원을 받아 수행되었음.

인식하는 데이터 이질성[4]을 보여준다.

는 메타데이터의 이질성 문제를 해결한다.

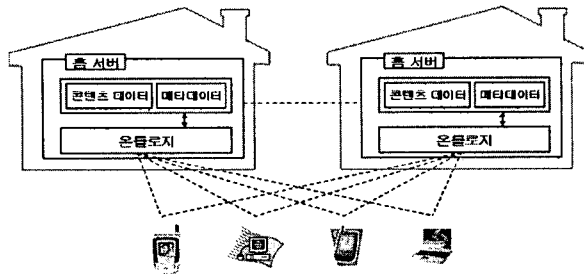


[그림 2] 홈서버와 모바일 단말의 연결 구성

기존 키워드 중심의 검색 방법으로 멀티미디어 정보에 대한 검색이 가능하지만 홈서버 내의 모든 메타데이터 요소를 비교, 검색해야 하므로 비용이나 시간적 측면에서 매우 비효율적이다. 하지만 메타데이터 요소간의 일관성이 유지된다면, 검색 시 동일한 멀티미디어 메타데이터만 검색하면 되므로 매우 효율적이다. 'title'이라는 메타데이터 요소에 '소풍'을 포함한 모든 사전을 검색한다고 가정하자. 기존 키워드 검색의 경우는 홈서버의 모든 메타데이터를 비교 검색하여 키워드 '소풍'을 포함하는 모든 결과를 보여준다. 반면, 메타데이터 요소의 일관성이 유지되면, 'title'와 관련된 메타데이터 요소를 비교, 검색하므로 더욱 용이하다. 따라서 효율적인 검색을 위한 각 메타데이터 요소간의 일관성 유지를 가능하게 하는 온톨로지가 필요하다. 본 논문에서는 메타데이터 간의 일관성 및 검색 효율을 증진시키기 위하여 메타데이터를 기반으로 하는 온톨로지 검색 방식을 제안한다.

3. 온톨로지를 적용한 메타데이터 기반 검색 설계

[그림 3]은 메타데이터와 온톨로지를 기반으로 하는 개선된 홈서버의 개요도이다. 개선된 홈서버는 콘텐츠 데이터에 대한 메타데이터를 가지며 온톨로지를 통하여 다른 홈서버와 단말 기기의 검색이 데이터의 이질성 없이 가능하다.



[그림 3] 메타데이터와 온톨로지 기반의 홈서버

3.1 메타데이터 관리를 위한 설계

멀티미디어 콘텐츠에 대한 메타데이터 요소를 정의하고 있는 MPEG-7 표준을 사용하며, [표 1]과 같이 홈서버 환경에서 보다 효과적으로 사용하기 위해 몇 가지 멀티미디어 메타데이터 요소를 추가, 변경했다.

메타데이터의 일관성을 유지하기 위해서는 [표 2]와 같은 메타데이터 요소의 속성을 먼저 정의해야 한다. 정의된 속성은 홈서버와 모바일 단말 등의 디바이스에서 메타데이터 요소의 사용자 입력을 제한한다. 따라서 각 디바이스마다 다를 수 있

[표 1] MPEG-7의 멀티미디어 메타데이터 구성 요소

데이터 요소	정 의
Title	리소스에 부여된 제목
Creator	리소스 콘텐츠를 만든 일차적 책임을 질 수 있는 존재
Subject	리소스 콘텐츠의 주제로 동영상, 사진, 음악 형식을 구분
Description	리소스 콘텐츠에 대한 설명
Publisher	리소스 콘텐츠를 만든 것에 대해 책임질 수 있는 존재
Contribution	공동 작업한 리소스 콘텐츠를 만든 것에 책임자
Date	리소스 생성주기에서 이벤트와 연관된 날짜로 보통 생성 날짜
Type	리소스 콘텐츠의 장르, 성질로 카테고리라 유사함
Format	파일 포맷 또는 타임 (MPEG-1, RealVideo, QuickTime)
Identifier	주어진 콘텐츠에서 리소스에게 주어지는 동일한 식별자
Source	파생된 현재 리소스로부터의 참조 정보
Language	리소스 콘텐츠에서 사용언어 정의
Relation	연관된 리소스에 대한 참조 정보
Duration	리소스 콘텐츠의 사용 영역 또는 확장 정보
Rights	리소스 소유에 대한 저작권 정보
Privacy	리소스를 타인에게 공개 또는 비공개
Grade	등급

[표 2] 메타데이터 요소 속성 정의

데이터 요소	속 성	데이터 요소	속 성
Title	String	Identifier	i0,i1,i2,...
Creator	room # 000	Source	String
Subject	String	Language	kor, Jpn ...
Description	String	Relation	String
Publisher	room # 000	Duration	yyyy.mm.dd
Contribution	String	Rights	room # 000
Date	yyyy.mm.dd	Privacy	String
Type	t0,t1,t2, ...	Grade	String
Format	f0,f1,f2, ...		

이렇게 정의된 메타데이터 요소들은 메타데이터 레지스트리 모델의 데이터 요소 개념으로 정립되기 위해서, 개념적 도메인, 값 도메인과 데이터 요소의 3가지 구성요소인 객체분류, 특성, 표현을 사용하여 구체화한다.

3.2 메타데이터 기반의 온톨로지 설계

각 홈서버는 메타데이터를 기반으로 하는 온톨로지를 가진다. 홈서버-모바일단말, 홈서버-홈서버 연결구조가 존재하며, 이들 사이에서 멀티미디어 콘텐츠의 검색 및 변경과 공유가 발생한다. 콘텐츠를 공유하기 위해서는 검색이 우선적으로 선행되어야 한다. 검색은 각 홈서버의 메타데이터 요소를 이용하지만, 이들 간에 일관성이 없다면 서로 다른 결과를 보여줄 수 있다. 따라서 멀티미디어 메타데이터의 요소들이 일관성 유지를 위해 온톨로지를 이용하여 공유될 수 있도록 한다.

정의된 메타데이터 요소들은 워드넷 형식의 온톨로지로 구성된다. 워드넷은 단어간의 관계를 표현하는 어휘 데이터베이스로 의미 중심으로 표현이 가능하다. 워드넷에서 사용되는 synset은 비슷한 의미를 가지는 단어들의 집합으로 링크를 통하여 synset들간의 의미적인 관계를 표현한다. 의미적으로 유사하다는 것은 문맥에서 단어들 이 서로 대체되었을 때, 의미적으로 큰 차이가 없다는 것을 의미한다. 가령, {부모, 부모님, 아버지}가 synset의 예이다. 워드넷 내에는 단어간의 동어어 관계, 반의어 관계 그리고 의미간의 상하위 관계, 부분-전체 관계 등이 있다. 하나의 synset에 있는 단어들은 동의어 관계에 있다. 단어간의 동의어 관계와 의미간의 상하위 관계가 시

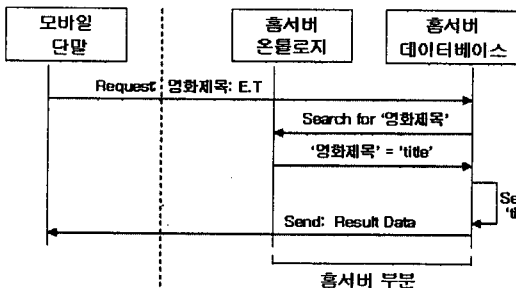
소스어를 이루는 기본 관계이다. 본 논문에서는 워드넷의 개념인 synset을 이용하여 MPEG-7에서 정의된 메타데이터와 유사한 단어들의 집합을 구성하여 검색에 사용한다.

[표 3]는 워드넷 형식으로 메타데이터의 요소들간의 동의어 정보를 표현한 것이다. 이와 같이 각 서버 내내 메타데이터 요소간의 동의어 정보를 저장해 두면, 서버-모바일 단말간의 데이터 검색 시 서로 다른 이름의 메타데이터 요소 이름을 사용한다고 하더라도 쉽게 메타데이터 요소간의 유사성을 파악한다.

[표 3] 메타데이터 요소간의 동의어 정보

Synet	Synonyms
Title	제목, 이름, 타이틀, subject, theme, heading, ...
Creator	생성자, 만든이, maker, person, name, ...
.....
Format	형식, 확장자, format, formatting, data format, ...

예를 들면, 모바일 단말에서 서버에 '영화 제목'이라는 메타데이터 요소에 'E.T'를 포함한 모든 영화를 요청한다고 가정한다. 서버는 단말로부터 검색 요청 정보를 받은 후 먼저 요청한 메타데이터 요소와 동일한 것이 있는지를 메타데이터 요소간의 동의어 정보를 이용해 파악한다. '영화 제목'이 'Title'과 동일한 데이터 요소라는 것이 파악되면, 서버는 자신의 데이터베이스에서 해당 콘텐츠를 검색하여 결과를 전송한다. 메타데이터 동의어 정보를 이용한 예시에 대한 검색 동작 과정은 [그림 4]와 같다.



[그림 4] 동의어 정보를 이용한 검색 동작 과정

서버와 단말간의 무선망을 통한 검색시 [표 4]와 같은 메타데이터 요소간의 동의어 정보가 없을 경우 키워드 중심의 검색을 하게 되며, 전체 시스템을 모두 검색하게 되어 시간 및 속도면에서 비효율적이다. 시간이나 속도적인 측면에서 매우 비효율적이다. 서버에 미리 저장된 메타데이터 요소 동의어 정보 용이하게 메타데이터 요소들간의 관계를 파악하여 보다 정확한 결과를 찾고 시간 및 속도면에서 빠른 장점을 가진다.

4. 평가

기존 서버 환경은 단말기를 통한 서버의 정보 검색시 뚜렷한 검색 방법에 대한 해결책이 미비한 상태이다. 서버는 멀티미디어 정보를 다수 개를 저장하고 있으며, 단말 사용자에게 무선망을 통한 정보를 제공해야한다. 따라서 검색에 필요한 시간이 길거나, 단말기와 서버간의 검색을 위한 패킷이 많이 발생하게 되면 불필요한 비용이 발생하게 된다. 본 논문에서 제시한 메타데이터와 온톨로지를 적용한 서버 검색 방법은 유사어를 통한 관련 데이터의 검색이 용이하며, 의미적 메타데이터의 일관성을 유지하고 있다. 따라서 콘텐츠 관리의 효율성

도 기존 서버에 비해 높다. 또한 일반 검색시 발생할 수 있는 검색실패를 방지하여 사용자에게 필요한 정보를 제공할 수 있다.

[표 4] 기존 서버, 메타데이터와 온톨로지 적용 환경 성능 비교

비교 항목	기존 서버	메타데이터 환경	온톨로지 환경
검색 용이성	용이하지 않음	비교적 용이함	용이함
의미적 메타데이터 일관성	없음	일부만 존재함	일관적임
콘텐츠 관리 효율성	낮음	중간	높음

5. 결론 및 향후 연구 과제

본 논문에서는 다수의 멀티미디어 정보를 저장하고 있는 서버를 무선망을 통하여 사용자에게 단말기를 통하여 데이터의 일관성 및 관련성을 유지하여 검색할 수 있는 온톨로지 기반 메타데이터 검색 방법을 설계하였다. 이를 위하여 서버에 저장된 멀티미디어 정보 공유를 위하여 데이터의 의미, 구문, 표현을 표준화하는 프레임워크인 메타데이터 레지스트리를 이용하여 멀티미디어 콘텐츠에 대한 메타데이터 요소를 정의하고 있는 MPEG-7 표준을 제정하였다. MPEG-7의 사용은 멀티미디어 콘텐츠의 메타데이터 요소들에 대한 의미를 부여하고, 멀티미디어 콘텐츠가 가지는 특징을 효율적으로 공유하도록 일관성을 제공한다. 또한 각 서버가 가지는 메타데이터를 기반으로 하는 온톨로지는 워드넷 형식으로 제공함으로써 서버와 모바일 단말간, 서버와 서버간의 콘텐츠 검색에서 기존의 키워드 중심의 검색에 비하여 검색의 용이성, 의미적 메타데이터의 일관성과 효율성의 측면에서 보다 월등한 성능을 보여준다.

향후 연구 과제는 서버 환경을 멀티미디어 콘텐츠 유통을 위한 프레임워크 표준인 MPEG-21로 확대 적용하는 것이다. DI (Digital Item)은 MPEG-21 프레임워크 내에서 표준화된 표현 형식, 식별체계, 서술 형식을 따르는 구조화된 디지털 객체이며, 유통, 처리의 최소 단위이다. 따라서 이러한 DI가 MPEG-21 멀티미디어 프레임워크 환경에서 사용자 터미널에 전달되었을 때 어떻게 처리되어야 할 것인지를 규정하는 것은 매우 중요한 과제이며, 이와 관련한 기술이 DIP(Digital Item Processing)이다. MPEG-21이 적용될 때 비로소 서버에서 이용되는 멀티미디어 콘텐츠에 대한 활용이 보다 완벽해 질 것이다.

6. 참고문헌

[1] 서광현, "디지털홍 구축 정책방향," TTA 저널, Vol.88, pp.20-29, 2003.8
 [2] 전호인, "디지털홍 기술 및 표준화 동향", 한국정보통신기술협회, TTA 저널, Vol. 88호, pp. 59-74, 2003.8
 [3] 류덕열 외, "디지털홍 상호운영성 기술", 한국정보통신기술협회, TTA 저널, Vol. 88호, pp. 92-98, 2003.8
 [4] 김진관 외, "메타데이터 이질성 해결을 위한 MDR기반의 메시지 변환 시스템", 한국정보과학회, 논문지D, Vol.31, No. 3, pp 0232-0242. 2004.06
 [5] 정동원 외, "데이터 가시성을 기반하는 점진적인 메타데이터 구축 방법론", 한국정보과학회, 논문지C, Vol. 09, No. 6, pp. 610-622, 2003.12
 [6] ISO/IEC JTC1/SC 32, ISO/IEC 11179: Specification and Standardization of Data Elements ISO/IEC JTC1, Part1-6
 [7] ISO/IEC JTC 1/SC 29, ISO/IEC 15938: Multimedia content description interface, Part1-4