

# 이질적인 멀티미디어 메타데이터 통합 관리를 위한 시스템 구조\*

박성규, 이양구, 채덕진, 류근호  
충북대학교 데이터베이스 연구실

e-mail : {parksk, leeyangkoo, djchai2520, khryu}@dblab.cbnu.ac.kr

## Integrated System for Managing Heterogeneous Multimedia Metadata

Seong Kyu Park, Yang Koo Lee, Duck Jin Chai, Keun Ho Ryu  
Database/Bioinformatics Laboratory, Chungbuk National University

### 요 약

최근 인터넷과 네트워크 기술이 급격히 발전함에 따라 대용량의 동영상 데이터에 대한 서비스 요구가 급격히 증가하고 있으며, 인터넷 사용자는 비디오, 오디오, 텍스트 등으로 표현되는 다양한 멀티미디어 콘텐츠를 쉽게 이용할 수 있게 되었다. 이러한 환경에서 동영상 콘텐츠의 저장, 교환 및 검색에 있어서 보다 나은 편리성을 요구하게 되었고, 이것은 동영상 데이터의 다양한 정보를 기술하기 위한 메타데이터의 중요성을 더욱 증대시키게 되었다. 그러나 이러한 서비스들은 데이터 포맷 및 서비스 종류에 따라 각기 다른 표준을 사용하고 있어 업계간 또는 서비스간에 다르게 표현되는 메타데이터 정보들을 통합 관리하기는 어려운 실정이다. 따라서, 이 논문에서는 기존의 메타데이터 관리 시스템의 이점을 수용하고, 이를 확장하여 각각의 표준에 따라 서로 다르게 표현되는 동영상 메타데이터를 통합 관리하기 위한 시스템을 제안한다.

### 1. 서론

오늘날 컴퓨터 처리 능력의 향상과 인터넷 기술의 급격한 발전은 대용량의 멀티미디어 데이터 처리를 가능하게 하였다. 이러한 환경에서 사용자들은 동영상 콘텐츠를 저장, 교환 및 검색함에 있어서 보다 나은 편리성을 요구하게 되었고, 이에 따라 동영상 콘텐츠의 다양한 정보를 논리적이고, 의미론적으로 구조화하여 표현하기 위한 메타데이터의 중요성이 더욱 증가하게 되었다[1].

그 동안 동영상 데이터를 보다 효율적으로 표현하고 검색하기 위해 많은 표준화 단체에서 메타데이터에 대한 표준화 연구를 진행하여 왔으며, 대표적인 연구로 더블린 코어(Dublin Core), TV-Anytime, ADI (Asset Distribution Interface), MPEG-7 등이 제안되었다. 그러나 이러한 메타데이터 표준들은 각각의 응용에 적합하도록 만들어졌기 때문에 오히려 메타데이터 표준들의 난립으로 인한 데이터 이질성 문제를 발생시킨다. 이러한 문제를 극복하기 위해 많은 연구에서는 이질적인 메타데이터 표준들을 통합 관리함으로써 정보 자원의 상호 운용성을 확보하는데 노력하고 있다. 그러나 이러한 노력은 특정 응용 분야에 제한되어 연

구됨으로써 사용자의 검색 욕구를 특정 서비스 영역으로 제한하는 문제가 있다.

따라서, 이 논문에서는 특정 서비스에 관계없이 동영상 정보를 검색 할 수 있도록 이질적인 메타데이터를 통합하고 이를 관리하기 위한 시스템을 설계하고 구현 한다. 제안된 시스템에서 각 서비스 간의 메타데이터 통합은 지금까지 제안된 메타데이터 표준들을 분석하고 각 표준 간의 공통 요소를 추출하여 통합 스키마를 구성함으로써 이루어진다. 또한, 일부 표준에 특화되어 독립적으로 구성된 이질적인 메타데이터 요소들은 각 표준이 제시하는 구조를 그대로 수용하여 통합 스키마에 추가된다. 이렇게 구성된 메타데이터는 미리 구축된 통합 스키마 구조에 따라 통합 저장됨으로써, 사용자는 찾고자 하는 동영상 정보를 다양한 서비스로부터 획득할 수 있다.

이 논문의 구성은 다음과 같이 구성된다. 2 장에서는 관련 연구를 통하여 기존의 메타데이터 표준을 소개하고, 기존 연구들에 대해서 설명한다. 3 장에서는 이 논문에서 제안된 시스템의 구조에 대해 기술한다. 4 장에서는 제안된 시스템의 구현 결과를 보인다. 마지막으로, 6 장에서는 결론과 향후 연구에 대해 기술한다.

### 2. 관련 연구

그 동안 동영상 데이터를 보다 효율적으로 표현하

\* 본 연구는 건설교통부 첨단도시기술개발사업 - 지능형국토정보 기술혁신 사업과제의 연구비지원(07KLSG02#) 그리고 교육과학기술부와 한국산업기술평산단의 지역혁신인력양성사업으로 수행된 연구결과임.

고 검색하기 위해 많은 표준화 단체에서 메타데이터에 대한 표준화 연구를 진행하여 왔다. 이러한 동영상 메타데이터를 위한 대표적인 표준으로 더블린 코어(Dublin Core), TV-Anytime, MPEG-7, ADI (Asset Distribution Interface) 등이 제안되었다[2, 3, 5].

더블린 코어는 네트워크 환경에서 정보를 쉽게 공유하고, 검색하기 위해 각 정보를 기술하는 핵심적인 데이터 요소를 규정한 표준으로 RDF(Resource Description Framework)를 기반으로 XML 을 사용하여 표현한다. MPEG-7 은 멀티미디어 콘텐츠에 관한 메타데이터를 기술하기 위해 2001 년에 제안된 표준이고, TV-Anytime 은 1999 년 미국에서 설립된 TV-Anytime 포럼에서 제정한 디지털 방송 관련 표준이다. 또한, ADI 는 케이블 산업과 고객들의 정보 관리를 위해 Cable Television Laboratories 회사에서 정의한 VOD 메타데이터이다.

동영상의 디지털 방송분야에서는 메타데이터의 특성을 고려하여 이를 효과적으로 저장하고 검색하기 위한 연구가 진행되었다. 이 연구에서 제안하고 있는 방법은 TV-Anytime 메타데이터 문서의 크기가 크고 문서의 일부분이 주로 재사용된다는 특성을 고려하여, 메타데이터를 단편으로 나누어 저장하고 검색하는 방법을 제안하고 있다. 또한, MPEG-A 에서 제안하는 MAF(Multimedia Application File Format)는 MPEG-7 형태의 메타데이터 특성을 갖고 있다. [9]는 이러한 특성을 이용하여 멀티미디어 데이터의 생성(Creation), 교환(Exchange), 변경(Modification), 검색(Search), 재생(Play)의 기능을 갖는 시스템을 구현하였다.

메타데이터 관리에 있어 중요한 문제의 하나는 서로 다른 출처에 있는 데이터들의 형식으로 인해 각각의 메타데이터들이 서로 조화되지 않는다는 것이다. 이러한 문제는 메타데이터의 구조적인 이질성뿐만 아니라 의미론적 이질성을 가중시키는 원인이 된다[4].

이 논문에서는 이러한 다른 분야에서 이용되고 있는 이질적인 동영상 메타데이터를 해결하기 위해 크게 두 가지 방법으로 접근한다. 첫 번째는, 다양한 표준들에서 공통 속성들을 추출하고 공통 속성들 중에서 대표되는 하나의 엘리먼트에 대해 선언하고 세분화한다. 두 번째는, 매핑 테이블을 이용하여 데이터의 의미는 같으나 스키마 속성 이름이 다른 경우와 의미와 속성 이름이 다른 경우에 대한 문제를 해결한다.

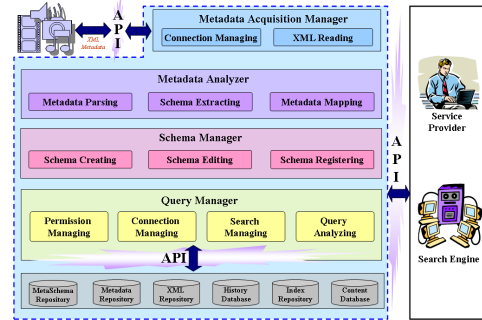
### 3. 제안된 시스템

#### 3-1. 시스템 구조

이 논문에서 제안하는 동영상 메타데이터 통합 관리 시스템은 크게 Metadata Acquisition Manager, Metadata Analyzer, Schema Manager, Query Manager 등의 모듈로 구성된다.

(그림 1)은 전체 시스템 구조를 나타내고 있다. 각각의 모듈에서 MetaData Acquisition Manager 는 외부로부터 메타데이터와 컨텐츠 데이터를 수신하는 기능을 수행한다. Metadata Analyzer 는 수신된 메타데이터를 파서를 이용하여 파싱하고 이를 분석하여 엘리먼트

속성과 속성값들을 추출한다. Schema Manager 는 추출된 속성 리스트들을 사전 정의된 스키마와 매핑하여 공통 속성 리스트를 생성한다. 마지막으로 Query Manager 는 매핑된 속성 리스트들을 데이터베이스에 저장한다.



(그림 1) 전체 시스템 구조

#### 3-2. 메타데이터 매핑

메타데이터의 매핑은 각 표준간의 의미적 관계를 분석하여 통합 스키마 구조를 정의하고 이에 따라 생성된 매핑 테이블을 통해 이루어진다. <표 1>은 생성된 매핑 테이블을 나타내는 예제이다.

<표 1> 매핑 테이블 스키마

한글 이름	영문 이름	설 명
속성이름	attributeName	본 시스템에서 정의한 스키마 속성이름
속성설명	attributeDescription	스키마 속성 설명
동속성	matchedAttribute	정의한 속성과 같은 내용을 갖는 속성
동속성값	matchedAttributeValue	동속성이 가지고 있는 속성 값
속성표준	attributeStandard	동속성이 정의된 표준의 이름
속성인덱스	attributeIndex	내용이 같은 다른 표준의 속성이 있을 때 그 속성에 비교된 정의한 속성의 인덱스

```

<Mpeg7>
  <Description>
    <MultimediaContent>
      <Video>
        <CreationInformation>
          <Creation>
            <Title>any</Title>
            <Creator>any<Creator>
            <PackagedType>any</PackagedType>
            <CreationDescription>any
            <CreationDescription>
              <date>any</date>
              <Genre>any</Genre>
              <FileFormat>any</FileFormat>
            </Creation>
          </CreationInformation>
        </Video>
      </MultimediaContent>
    </Description>
  </Mpeg7>
    
```

(그림 2)Mpeg-7 메타데이터

```

<TVMain>
  <MContentDescription>
    <MProgramInformationTable>
      <ProgramInformation>
        <BasicDescription>
          <ShortTitle>any</ShortTitle>
          <CreditsItem>any</CreditsItem>
          <Synopsis>any</Synopsis>
        </BasicDescription>
      </ProgramInformation>
    </MProgramInformationTable>
  </MContentDescription>
</TVMain>
    
```

(그림 3)TV-Anytime 메타데이터

(그림 2)는 MPEG-7 의 메타데이터의 예제이고, (그림 3) TV-Anytime 의 메타데이터 예제이다.. 각 표준마다 더 많은 속성이 있지만 간단히 구조만을 표현하였다. <표 2>는 (그림 2)와 (그림 3)에서 각 표준들의 공통 속성을 분석한 것이다. <표 2>에서 볼 수 있듯이 각 표준마다 같은 의미를 갖는 속성들을 다르게 표현하는 것을 알 수 있다.

<표 2> Dublin Core, MPEG-7, TV-Anytime 공통속성

Dublin Core	MPEG-7	TV-Anytime
Title	Title	Title, ShortTitle
Creator	Creator	CreditsItem
Subject	PackagedType	Synopsis
Description	CreationDescription	ProgramDescription
Date	Date	CreationDate
Type	Genre	Genre
Format	FileFormat	FileFormat

**Algorithm ExtractSchema (imeta)**

```

Input : imeta (메타데이터 인스턴스)
Output : schemaListSet (스키마 리스트 인스턴스)
Begin
  ptree = parseXML(imeta)
  proot = getRootElement(ptree)
  while(count of node in proot)
    ename = getElementName
    (proot.element)
    icontent =
    getElementValue(proot.value)
    schemaListSet = addList(ename,
    icontent)
    proot = getNextElement(proot)
  end while
    
```

```

return schemaListSet
end
    
```

(그림 4) 스키마 리스트 추출 알고리즘

**Algorithm MappingSchema (schemaListSet)**

```

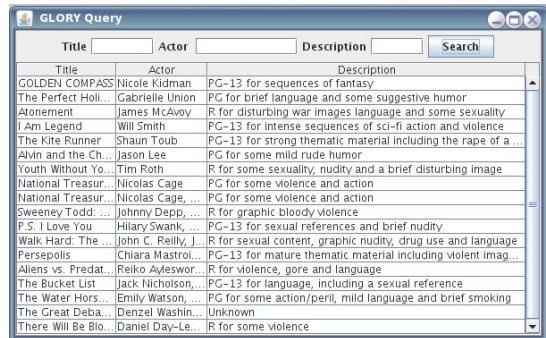
Input : schemaListSet (스키마 리스트 인스턴스)
Output : mappedListSet (매핑된 스키마 리스트 인스턴스)
Begin
  mappedListSet = schemaListSet
  dbCon = connect database
  while(count of list in schemaListSet)
    temp = find schema name the same
    with schemaListSet.name in
    mapping table
    mappedListSet.name = temp
  end while
  return mappedListSet
end
    
```

(그림 5) 스키마 매핑 알고리즘

(그림 4)는 메타데이터 분석에서 메타데이터 파싱을 통해 스키마 리스트를 추출하는 과정을 알고리즘으로 나타낸 것이며, (그림 5)는 추출된 스키마 리스트를 매핑 테이블을 이용하여 통합 스키마로 매핑하는 알고리즘을 나타낸 것이다. 이러한 매핑 알고리즘을 통하여 메타데이터의 표준에 관계없이 원하는 결과를 얻을 수가 있다.

**4. 프로토타입 구현**

제안된 시스템은 Intel 2.4GHz, RAM 1GB 환경에 운영체제로는 리눅스를 사용하여 구현하였고, 개발 언어로는 JAVA 를 사용하였다. 또한, XML 포맷의 메타데이터 파일을 파싱하기 위해 SAX2, DOM 파서를 이용하였고, 데이터베이스는 My-SQL 5.0 을 이용하여 구축하였다. 입력된 데이터는 TV-Anytime, MPEG-7 의 타이틀 메타데이터를 이용하였고, 사용자 질의에 대한 검색은 통합된 매핑 테이블을 통해 이루어진다.



(그림 6) 매핑 테이블의 전체 데이터

Title	Actor	Description
The Perfect Holiday	Gabrielle Union	PG for brief language and some suggestive humor
The Kite Runner	Shawn Toub	PG-13 for strong thematic material including the rape of a ...
The Bucket List	Jack Nicholson	PG-13 for language, including a sexual reference
The Water Horse...	Emily Watson	PG for some action/peril, mild language and brief smoking
The Great Debat...	Denzel Washing...	Unknown
There Will Be Blood	Daniel Day-Lew...	R for some violence

(그림 7) Title 요소에 대한 질의 처리 결과

(그림 6)는 매핑 처리된 후 미리 정해진 매핑 테이블에 저장되어있는 모든 데이터를 보여주고 있다. (그림 7)은 Title 요소를 이용하여 질의한 결과를 보여 주고 있다. (그림 7)에서 볼 수 있듯이, 통합된 매핑테이블을 통해 각 서비스 표준에 관계없이 검색된 결과가 나타나는 것을 볼 수 있다.

**5. 결론 및 향후 연구**

이 논문에서는 동영상 데이터 표현의 불일치 및 이질성 문제와 각 분야의 메타데이터간의 상호 운용성 문제를 해결하기 위한 통합 메타데이터 관리 시스템을 설계하고 프로토타입 시스템을 구현하였다. 제안된 시스템은 더블린 코어(Doublin Core), MPEG-7, TV-Anytime 표준에서 제정한 메타데이터의 분석을 통하여 어의적 매핑 후 공통 스키마를 추출하여 통합 스키마를 구성 하였고, 이렇게 구성된 메타데이터는 미리 구축된 통합 스키마 구조에 따라 매핑하여, 사용자가 찾고자 하는 동영상 정보를 응용 서비스에 제한 받지 않고 획득 할 수 있게 하였다.

향후 연구로써, 이 논문에서 제한적으로 사용한 각 동영상 표준들에 대한 분석 범위와 적용 범위의 확장과 제안된 시스템의 성능 평가가 필요하다.

**참고문헌**

[1] 송치일, 낭종호, “MPEG-7 시각 정보 기술자의 인덱싱 및 결합 알고리즘”, 정보과학회논문지, 34 권, 1 호, pp. 1-10, 2007.

[2] 강경욱, “TV-AnyTime 포럼”, TTA 저널, 제 80 호, 2002.

[3] 노승민, 황인준, “멀티미디어 검색 시스템의 설계 및 구현”, 정보과학회논문지, 30 권, 5 호, pp.494-506, 2003.

[4] 류성준, 황재문, 김태훈, 남영광, “메타정보 인터페이스를 이용한 이질 구조 분석 XML 문서

통합 검색”, 정보과학회논문지, 31 권, 11 호, pp.1505-1518, 2004.

[5] 이경하 외, “XMF: XML 기반 분산 이질 정보 자원의 통합 프레임워크”, KDBC2000, pp. 262-270, 2000.

[6] 신효섭, “디지털 TV 방송 환경에서 내장형 시스템을 위한 XML 데이터의 저장 및 검색 방법”, 데이터베이스연구, 19 권, 3 호, pp.58-66 2003.

[7] B. M. Heo, " DESIGN OF METADATA MANAGEMENT SYSTEM FOR RETRIEVAL OF VIDEO DATA", In: ISRS 2007 International Symposium on Remote Sensing, Korea, pp. 314-316, 2007.

[8] Gnther O, Voisard A, "Metadata in geographic and environment data management", in Sheth A, Klas W (eds) Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media. McGraw Hill: 57-87. 19 98.

[9] Y.M Gang, “A design and Implementation of Multimedia Retrieval System on MAF(Multimedia Application File Format)”, In: The Korean Institute of Information Scientists and Engineers(KIISE), Vol.33, part 9,2006.

[10] Ram, S. and Ramesh, V. Schema Integration: Past, Current and Future, in A. Elmagarmid, M. Rusinkiewica, and A. Sheith, editors, Management of Heterogeneous and Autonomous Database System, *Mogen Kaufmann Publishers*, pp.119-155, 1999.