

XTM을 이용한 MDR기반 콘텐츠 메타파일 관리 시스템 설계

유우종**, 임희영*, 임정은*

*고려대학교 컴퓨터학과 소프트웨어 시스템 연구실

e-mail:{milkybell**, tonight24, jelim}@software.korea.ac.kr

Design of a MDR based Contents Metafile Management System using the XTM

Woo-Jong Yoo^o Hee-Young Lim Jung-Eun Lim

Software System Lab. Department of Computer Science and Engineering, Korea University

요 약

콘텐츠 관리 시스템(CMS:Contents Management System)은 ‘자료 수집, 등록, 검색, 배포’의 기본 흐름을 가진다. 콘텐츠의 등록 및 검색/배포를 위하여 각 시스템은 콘텐츠에 대한 별도의 메타파일들을 가지고 있으나, 이러한 메타파일들은 데이터 요소의 중의성이나 모호함 때문에 일관되고 객관화 된 스키마를 가지지 못하여 체계적 분류 및 최신 업데이트를 위한 메타데이터 자체의 효율적 관리 및 연관 검색 기능을 가지고 있지 않았다. 본 논문에서는 기존 연구되고 있는 MDR과 토픽맵을 자체 개발 중인 콘텐츠 메타파일 관리 시스템(CMMS:Contents Metafile Management System)에 적용하여 메타파일의 체계적이고 효율적인 관리를 통해 기준요소로서의 메타파일 역할을 극대화하고 향후 타 체계와의 연동 및 확장성의 향상을 도모한다.

1. 서 론

최근 급속도로 증가하는 멀티미디어 콘텐츠에 대한 Raw Data의 수집과 배포의 수요가 증가하고 있으며, 그에 따른 다양한 콘텐츠 관리 시스템 (CMS : Contents Management System, 이하 CMS)이 유통되고 있거나 필요에 따라 자체 제작 되어 사용되고 있다. CMS는 ‘자료 수집, 등록, 검색, 배포’의 기본 흐름을 가진다. 이 때, 콘텐츠의 등록/검색을 위하여 각 시스템은 콘텐츠에 대한 별도의 메타파일과 독특한 멀티미디어 검색 방식들을 가진다.

멀티미디어 데이터의 검색은 크게 내용 기반 검색과 주석 기반 검색으로 나눌 수 있다.

주석 기반 검색은 내용기반 검색에 비하여 상대적으로 제작비용이 적게 소요되지만, 멀티미디어 데이터에 대한 객관적이고 정확한 표현이 어렵기 때문에 모든 정보를 일관되게 제공하기 어렵다는 단점이 있다. 또한 주석 기반 검색은 1차원적인 검색 결과 즉, 사용자가 입력한 텍스트와 매칭하여 이를 포함하는 주석을 가진 콘텐츠를 리스트화 하는 방식의 결과를 보여준다. 사용자는 찾고자 하는 콘텐츠의 제목을 선택하여 주석내용을 확인하고 관심사항에 대해서는 그 내용을 직접 추출하여 검색 엔진을 통해 재검색하여야 하는 제약점이 있다.

MDR을 이용하여 메타데이터 요소에 대해 표준에 의거 해석할 수 있는 체계를 정립하고 의미 정보를 체계적으로 표현하기 위한 스키마와 토픽맵을 구축하여 주석 기반 방식의 장점인 경제성과 미묘한 의미 전달기능을 그대로 유지하면서 상가에서 기술한 2가지의 단점(객관성 결여, 연관검색 불가)을 보완한다면, 적은 비용으로 높은 정확도와 객관성을 가지는 효과적인 멀티미디어 검색이 되도록 할 수 있다.

본 논문에서는 기존에 연구되어 온 MDR과 토픽맵을 콘텐츠 메타파일 관리 시스템 (CMMS : Contents Metafile Management System, 이하 CMMS)에 적용하여 멀티미디어데이터 메타파일의 체계적이고 효율적인 관리를 통해 기준요소로서의

메타파일 역할을 극대화하고 향후 타 체계 연동 및 확장성의 향상을 도모하였다.

2장에서는 기존에 연구되어 온 멀티미디어 검색시스템과 MDR, 그리고 토픽맵에 대한 관련 연구를 알아보고, 3장에서는 CMMS의 주요 기능과 원리를 알아보고, 4장에서는 기존 방식과의 비교를 통해 주요 개선점을 평가하였다. 끝으로 5장에서는 이 논문 전체의 결론과 향후과제를 정리한다.

2. 관련 연구

2.1 멀티미디어 검색 시스템

최근까지 멀티미디어 검색 시스템과 관련된 여러 가지 프로토타입 시스템이 제안되었는데, Chabot[2] (복합 방식), VisualSEEK[3](내용 기반 방식), QBIC (주석 기반 방식), SMOOTH Video DB, VideoQ, Informedia, Virage 등의 많은 검색 시스템들이 있다. 여기서 사용되는 멀티미디어 데이터의 검색방식을 내용 기반 검색과 주석 기반 검색으로 나누어 살펴보면, 내용 기반 검색은 객체의 색상, 질감, 형태, 위치, 동작 등 시각적인 정보와 청각적인 정보를 기반으로 이루어지며, 주석 기반 검색은 데이터의 의미나 내용 등에 대해 서술한 정보를 바탕으로 이루어진다. 내용 기반 검색은 비용이 많이 들고 정확도가 낮은 단점이 있으며, 주석 기반 검색은 데이터가 커지고 종류가 다양해짐에 따라 주석 처리에 소요되는 시간이 증가하고 사람의 주관적인 판단에 따라 기술됨으로써 서술의 객관성에 문제점이 있다.

내용 기반 검색에 대한 최근 연구에서는 영상과 오디오 정보를 이용하여 좀더 효과적인 검출을 할 수 있는 기법들이 많이 선보이고 있다. 내용 기반 검색방식의 성능향상을 위한 연구가 계속적으로 이루어져 왔지만, Visual 데이터에 대해 쉽고 빠르게 의미 정보를 해석하는 것은 사람이며, 사람에 의한 주석처리는 멀티미디어 콘텐츠 관리에 있어서 필수적이라고 할 수 있다. 본 논문에서는 기본 검색방식으로 주석기반 검색방식을 이용하였다.[1]

2.2 MDR(Metadata Registry)

기존 시스템의 통합과, 여러 시스템 사이의 정보 공유에 대한 요구에 부응하기 위해 다양한 방법론이 개발되어 왔으나, 가장 큰 걸림들은 시스템적인 통합 메커니즘의 개발과 더불어, 의미(semantic), 구문(syntax), 표현(representation)의 불일치를 해결해야 하는 것이었다.

이에 데이터 표준화를 위한 작업의 일환으로 ISO/IEC JTC1 WG2 - 데이터 관리 및 교환(Data Management and Interchange)에서는 데이터 공유 및 교환을 위한 근본적인 해결 방안으로 메타데이터 레지스트리(Metadata Registry : MDR)에 대한 표준화를 진행하고 있다. 국제표준 ISO/IEC 11179 - “메타데이터 레지스트리(정보 기술 데이터 요소의 명세와 표준화)”에서 데이터의 의미, 구문, 표현을 표준화할 수 있는 프레임워크를 제시하고 있다.

메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 유지·관리하며, 메타데이터의 명세와 의미의 공유를 지원하고 있다.[4]

2.3 토픽맵(Topic Maps)

토픽맵은 정보 자원들을 상호 연관성에 따라 연결하고 조직화하여 지식 구조를 기술할 수 있도록 제정된 ISO 표준이다. 대용량의 정보를 분류하고 구조화하며 의미론적인 연관관계를 설정할 수 있는 모델을 제시하고 있으며 원하는 지식을 쉽고 정확하게 찾을 수 있는 맵을 제시한다.

토픽맵 모델은 토픽(Topic), 연관관계(Association), 어커런스(Occurrence)의 기본요소로 나누어진다.

- 토픽(Topic)

토픽은 일반적으로 어떠한 것이든 될 수 있다. 즉 사람, 사물, 개념, 의미 등 실제 존재하는 것, 또는 특정 속성이나 어떤 의미 등이 될 수 있다. 보통 명사형으로 표현된다. 비슷한 유형에 속하는 것끼리 분류를 해야 하는데 이때 토픽 타입이라는 것을 사용한다.

- 어커런스(Occurrence)

각 토픽은 자신이 참조하는 하나 이상의 실제적인 지식 항목(내용)과 연결될 수 있다. 예를 들어 “태극기 휘날리며”라는 토픽은 영화의 타이틀이고 이것은 “http://www.taeguki2004.com”라는 페이지를 가리키는 것으로 자원(Resource)과 연결될 수 있다. 이러한 연결정보를 어커런스라고 한다.

- 연관관계(Association)

토픽맵 표준안 스펙에는 토픽들 간의 연관관계를 정의할 수 있는 방법을 제공하고 있다. 토픽 연관 관계는 둘 이상의 토픽들 사이에 상하관계가 아닌 의미적인 관계성을 정의하는 관계이다. 토픽맵 내의 토픽들은 서로 독립적인 객체이면서 동시에 특정 타입의 연관관계로 연결된 링크를 가진다. 토픽들이 토픽 타입이라는 것으로 분류가 되는 것처럼 토픽들 간의 연관관계도 연관관계 타입이라는 것으로 분류된다. “made_by”, “play”, “broadcast_on” 등이 연관관계 타입이다.

토픽 타입과 연관관계 타입은 지식 및 정보 표현, 분류, 구조화를 위한 토픽맵의 중요한 기능이다. 직관적이면서 사용자에게 친숙한 인터페이스를 제공할 수 있을 뿐만 아니라 토픽맵을 네비게이션 할 때 사용자가 관심 있는 분야를 쉽게 축소할 수 있고 검색결과 또한 정확하게 만들어 낼 수 있다. 연관관계에는 타입 외에도 “역할(Role)”이라는 중요한 정보가 있다.[5]

3. CMMS

CMMS는 멀티미디어 콘텐츠의 관리 및 유통을 위한 표준 메타데이터 요소를 추출하여 MDR 요소의 등록 및 검색 기능을 수행하며, 이를 기반으로 생성된 스키마에 따라 사용자가 요청한 Contents의 메타정보를 등록, 관리하는 시스템이다. ‘그림1’은 CMMS의 전체 구성을 보여준다.

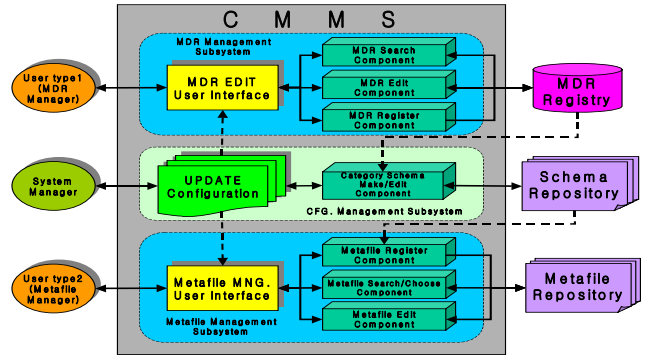


그림 1 CMMS 구성도

CMMS를 위 ‘그림1’에서 보여지는 서브 시스템 별로 설명하면 다음과 같다.

- 메타데이터 요소 관리(MDR Management)

메타데이터 관리는 멀티미디어 콘텐츠 관리 및 유통을 위해 추출된 데이터 요소를 MDR에 등록하고 검색하는 기능을 제공한다.

- 카테고리 스키마 관리(System Cfg. management)

메타파일 내의 데이터 요소를 콘텐츠의 분류별로 구성하기 위해 사용한다. MDR에 저장된 메타데이터를 이용하여 데이터 요소를 구성하고 스키마를 생성한다.

- 메타파일 관리(Metafile Management)

메타파일 등록기는 스키마에서 구성된 데이터 요소에 대한 값으로 콘텐츠 정보를 입력하여 메타파일을 생성하는 기능을 한다.

메타파일 검색기는 작성된 메타파일을 데이터 요소를 기준으로 검색하여 원하는 멀티미디어 콘텐츠를 찾는 기능을 제공한다. 검색 결과를 이용하여 콘텐츠 배포 등을 위한 의사결정 자료를 제공해 준다.

CMMS의 가장 핵심적인 부분은 표준 데이터 요소를 추출하여 MDR에 등록하는 것과 등록된 MDR 요소로 만들어진 메타파일 스키마로 XML을 이용한 토픽맵(XML Topic Maps 이하 XTM)을 작성하는 것이다.

3.1 MDR 데이터 요소 속성 스키마

먼저 MDR 등록을 위해 필요한 데이터 요소를 ‘ISO/IEC 11179’에 의거하여 추출하고, 각 요소에 대하여 속성/목록별로 명확하게 구분하여 명세한다. 이렇게 명세 되어진 데이터요소들은 MDR에 등록되어 카테고리 스키마 생성 및 메타파일 관리에서 타 시스템과 구분되는 표준요소로서의 기준 분류자 역할을 하게 된다.

MDR에서 정의하는 속성 중 멀티미디어 관리를 위하여 필요한 해당 속성을 추출하고 메타파일 생성을 위한 스키마를 작성한다. ‘그림2’는 멀티미디어 데이터 요소를 MDR에 등록하기 위해 작성한 데이터 요소 속성 스키마를 보여준다.

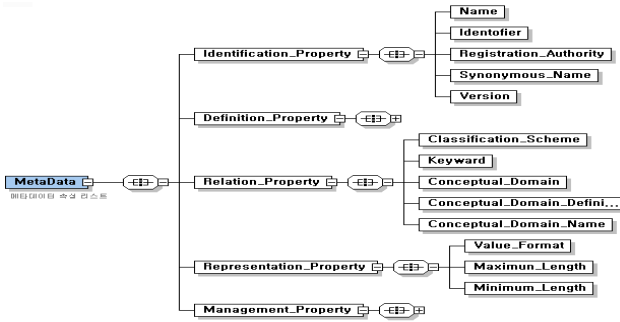


그림 2 MDR 데이터 요소 속성 스키마

MDR Manager는 멀티미디어 데이터 요소 검색 모듈에서 데이터 요소로 등록되어 있는 메타데이터들을 검색하고, 검색을 통해 자신이 원하는 데이터 요소를 찾을 수 없는 경우 원하는 데이터 요소를 MDR에 메타데이터로 등록한다.

3.2 메타파일 데이터 요소 스키마

사용자가 이러한 기능을 이용할 수 있도록 스키마를 만들어 제공해 주는 System Manager는 카테고리 스키마 편집 컴포넌트를 이용하여 2단계의 계층으로 나누는데, 상위계층은 ‘공통 요소’와 ‘카테고리별 세부요소’로 분류되어 지고, 하위계층은 ‘카테고리별 세부요소’에서 6개의 서로 다른 종류(영화, 드라마, 교양오락, 뉴스, 스포츠, 음악 등)로 분류된 각각의 스키마를 가진다. 본 시스템에서는 MDR에 등록된 멀티미디어 데이터의 공통요소로서 2002년 3월에 ISO/IEC JTC1/SC29 WG11에서 국제표준화 작업이 완료된 MPEG-7의 표준 데이터 요소를 반영하였다. ‘그림3’은 추출된 표준 데이터 요소를 이용하여 메타파일 등록을 위해 만들어진 멀티미디어 데이터 공통요소 스키마를 보여준다

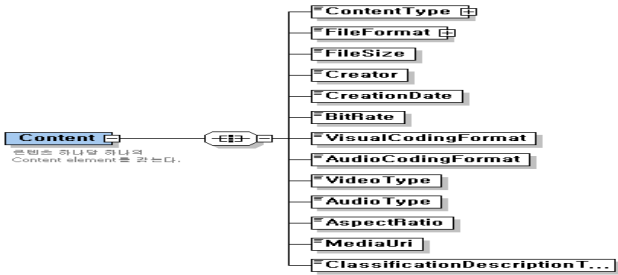


그림 3 멀티미디어 데이터 공통요소 스키마

3.3 XTM을 이용한 연관검색

CMMS의 중요한 특징 중 하나로 XTM을 이용한 메타파일 연관검색을 들 수 있는데, 그 기능의 구성과 원리는 다음과 같다.

먼저 MDR Manager는 앞으로 기준으로 쓰이게 될 필요한 데이터 요소들을 MDR에 등록한다.

MDR을 이용하여 본 연구팀이 기등록한 영화와 뉴스 스키마의 데이터요소는 표 1, 표 2와 같다. 여기서 MDR에 등록된 데이터요소 “Title”은 그 의미와 기능이 토픽맵의 “Topic Type”요소에 해당된다. 영화의 데이터요소들을 그 기능별로 각각 토픽맵의 요소로 매핑 시켜보면 표 1의 3번째 컬럼에 명세된 것과 같이 전환 될 수 있음을 알 수 있다. 4번째 컬럼은 MDR 기본속성 중 “개념영역”이라는 속성이고 그 값은 “Text, Person, Technique, Representation”

의 4개의 영역으로 나뉘어 지며 이 영역은 검색 개념적으로 서로 일치되고 포함되는 요소에 대해서만 접근/검색하여 역세스 타임과 튜닝 타임을 줄여 줄 수 있는 기능을 제공 해 준다. 예를 들어, 사용자가 감독에 해당하는 “강제규”를 검색하려고 할 경우, 검색기는 키워드 입력 폼에서 ‘Person’란에 입력된 “강제규”라는 단어를 읽어 들이고, 이를 MDR의 개념영역을 참조하여 해당 값이 ‘Person’인 요소만 검색하여 그 결과를 돌려주게 된다.

표 1 영화 스키마 데이터 요소

데이터요소	영 문	토픽맵 매핑	개념 영역
제목	Title	Topic Type	Text
제작자	Producer	Topic Type	Person
감독	Director	Topic Type	Person
출연	Actor	Topic Type	Person
원작	OriginTitle	Topic Type	Text
언어	Language	Topic Type	Representation
제작년도	PublicationDate	Topic Type	Technique
국가	Country	Topic Type	Text
음악	OST	Topic Type	Text
원제	FormerTitle	Topic Type	Text
줄거리	MediaReview	Topic Type	Text
장르	Genre	Topic Type	Text
카테고리	Category	Topic Type	Text
상영시간	Runningtime	Topic Type	Technique
관람등급	AudienceRange	Topic Type	Technique
회수	Serisee	Topic Type	Technique
위치	MediaUrl	Occurrence	Technique

표 2 뉴스 스키마 데이터 요소

데이터요소	영 문	토픽맵 매핑	개념 영역
뉴스제목	Newstitle	Topic Type	Text
기사제목	Articletitle	Topic Type	Text
방송날짜	BroadcastDate	Topic Type	Technique
앵커	Anchor	Topic Type	Person
기자	Journalist	Topic Type	Person
제작국	Producer	Topic Type	Text
카테고리	Category	Topic Type	Text
방영시간	RunningTime	Topic Type	Technique
위치	MediaUrl	Occurrence	Technique

이제 필요한 데이터 요소들이 모두 이미 등록되어 있다고 가정하면, 사용자가 콘텐츠에 대한 메타파일을 생성 할 때 이에 대응하는 토픽맵이 자동으로 생성된다. 그림 4와 그림 5는 “태극기 휘날리며”라는 영화와 그 영화에 관련된 기사가 포함된 뉴스에 대한 토픽맵이 생성된 모습을 간략히 보여준다.

이렇게 생성된 토픽맵을 사용자는 실제 검색에서 다음과 같은 용도로 사용 할 것이다.

- 사용자는 최근 흥행하는 영화와 그에 관련된 소식을 종합하여 미국에 있는 친척에게 보내려고 한다.

사용자는 먼저 “태극기 휘날리며”를 키워드로 검색하여 그에 따른 메타파일을 볼 수 있다. 이어서 사용자는 요소 값 중 “장동건”(“태극기 휘날리며”와 ‘play’라는 Association으로 연결된 ‘Topic’)을 선택하여 해당 홈페이지(Topic’장동건’의 Occurrence)를 볼 수 있다. 또 “태극기 휘날리며”(MDR 개념 영역 중 ‘Text’)를 선택하여 그와 관련된 뉴스인 “태극기 휘날리며, 900만 돌파”(MDR 개념영역 중 ‘Text’)라는 기사도 연관하여 검색 할 수 있다. 이 모든 것은 스키마에 의해 생성된 토픽맵과 MDR의 참조 기능으로 가능하다

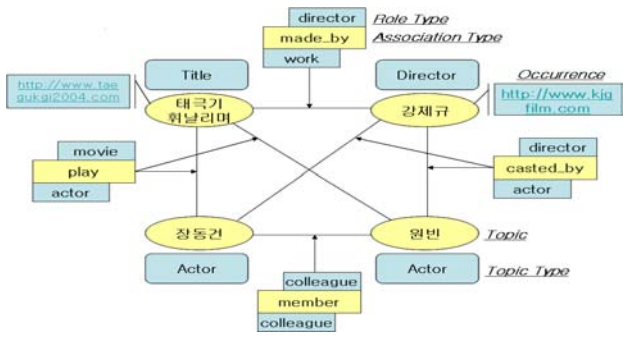


그림 4 영화 토픽맵 예

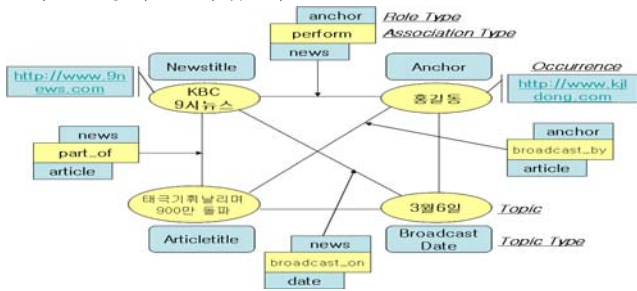


그림 5 뉴스 토픽맵 예

4. 비교 및 평가

본 논문에서 구축하는 MDR을 이용한 CMMS의 개선점을 설명하기 위해 다른 CMS에서 사용되는 검색방식과 비교/평가하였다.

복합방식 검색은 향후 가장 이상적인 검색 방식으로서 검색반응도(Recall)와 정확도(Precision)가 높다. 그러나, 현재는 낮은 경제성이 단점이다. 내용기반 검색은 주석으로는 감지 할 수 없는 콘텐츠의 특성을 자동으로 검색하고, 기계에 의존하기 때문에 객관성이 높다. 반면에, 기술구현의 한계성 때문에 찾고자 하는 데이터를 찾아내는 정확도가 다른 방식에 비해 현저히 떨어진다. 주석기반 검색은 구현이 간단하며, 기계가 인식하기 어려운 부분을 사람이 직접 이해하여 입력, 검색 할 수 있기 때문에 정확도가 비교적 높은 반면에, 동일한 이유에 의하여 객관성과 일관성이 매우 낮고, 처리소요시간이 과다하게 소요되는 단점을 가지고 있다. MDR을 이용한 주석기반 검색은 복합방식에 비하여 경제적이고, 내용기반검색에 비하여 정확도가 높으며, 주석기반검색에서 부족한 객관성을 높여줄 수 있을 뿐만 아니라 연관검색이 가능하다는 장점이 있다. '표3'은 비교내용을 표로 보여준다.

표 4 멀티미디어 검색 방식별 비교

구분	복합 방식 검색	내용기반 검색	주석기반 검색	MDR이용 주석기반검색
제작비용 경제성	· 매우 낮음	· 낮음	· 매우 높음	· 높음
정확도 (Precision)	· 높음	· 매우 낮음	· 높음	· 매우 높음
검색반응도 (Recall)	· 매우 높음	· 낮음	· 매우 낮음	· 보통
객관성 (일관성)	· 보통	· 매우 높음	· 낮음	· 높음
표준준수 여부	· 일부 준수	· 표준 없음	· 일부 준수	· 준수
연관검색 여부	· 불가 (일반적)	· 불가 (일반적)	· 불가	· 가능

본 시스템의 구체적인 개선사항을 보면,
 - 메타데이터 요소를 ISO/IEC 11179 표준을 준수하여 MDR에 사전 등록 후 사용함으로써 각 요소별 중의 성 및 호환성을 해소하였고,
 - 메타파일 생성 스키마를 규격화하여 데이터 요소 입력 시간을 줄였으며, 효율적인 메타파일 생성/검색이 되도록 하였다.
 - 다양한 분류자 및 관련사항 추가기능을 통해 파일의 형식에 무관하게 분류 및 검색이 가능하게 하였고,
 - XTM을 이용하여 연관검색이 가능하게 하였으며,
 - 다른 CMS나 단순 저장/검색 시스템에서도 표준화된 용어 및 데이터 관리를 위하여 생성해 놓은 MDR의 이용이 가능하여, 이를 통해 많은 양의 멀티미디어 데이터가 서로 공유하고, 호환할 수 있도록 표준화 된 환경을 제공해 준다.

5. 결론 및 향후 과제

본 논문에서는 기존 주석기반 검색방식에 MDR을 이용하여 멀티미디어 콘텐츠의 메타파일을 관리하는 방법을 제안하였다. 기존 시스템들은 멀티미디어 데이터에 대한 메타데이터 모델이나 스키마가 서로 다르고 교환의 표준이 정해지지 않아 통합하는데 어려움이 있다. 뿐만 아니라, 멀티미디어 데이터가 내포하는 다양하고 복잡한 정보를 효율적으로 처리하는데 제약점이 많다. 멀티미디어 콘텐츠의 이상적인 검색방법은 내용기반과 주석기반을 병용하는 복합 방식이겠지만, 아직 기계의 내용 인식 정도가 사람이 할 수 있는 미묘한 의미 인식에 미치지 못하는 것이 사실이다. 또한 복합방식은 제작비용이 많이 소요되기 때문에, 추가한 비용에 비하여 얻을 수 있는 정확도(Precision)와 검색반응도(Recall)는 상대적으로 낮다. 반면에, 주석기반 방식의 검색방법은 제작비용이 적게 소요되지만, 콘텐츠의 메타파일을 만들기 위한 주요 의사결정을 사람이 담당하기 때문에 객관성과 일관성에서 문제점이 제기된다.

이에, 본 논문에서는 국제표준 ISO/IEC 11179를 따르는 MDR로 데이터요소 DB를 구축하여, 이를 이용한 CMMS의 설계를 제안하였다. CMMS에서 이용하는 관리방식은 복합방식에 비하여 경제적이고, 내용기반방식에 비하여 정확도가 높으며, 주석기반방식의 가장 큰 문제점인 객관성과 일관성을 향상시켜 주었다. 게다가 주석기반방식의 1차원적인 결과값을 토픽맵 기술을 응용하여 관계성과 의미가 살아있는 입체적 검색방식으로 개선하였다. CMMS는 앞으로 검색반응도를 높이고 처리소요시간을 줄이기 위해서 내용기반 자동추출 기술의 적용이 필요하며, 주석기반 방식뿐만 아니라, 내용기반 방식의 데이터 요소들도 추출하여 MDR을 통한 표준화 등록이 필요 할 것이다. 향후 여러 연구들에서 응용 될 수 있는 표준화와 상호 호환성의 향상을 위한 방향을 제시한 데에 본 연구의 의의가 있다고 할 수 있다.

참고문헌

- [1] Y. Nam, E. Hwang "XMARS : XML-based Multimedia Annotation and Retrieval System" The KIPS Transactions : Part B, Vol.9-B, NO 5, October 2002.
- [2] Virginia E. Ogle and Michael Stonebraker, "Chabot"FRetrieval from a Relational Database of Images," IEEE Computer, Vol.28, No.9, pp.40-48, September, 1995.
- [3] J. R. Smith and S. F. Chang. "VisualSEEK"Fa fully automated content-based image query system," ACM Multimedia, Boston, May, 1996.
- [4] 한국과학기술정보연구원 "A study on the development of standardization and management model for science and technology information" 한국과학기술정보연구원, November, 2002.
- [5] H. Jeong, J. Kim, J.Jung, H. Kim "Knowledge Map of XTM Base" The Database Research : Vol.19, NO 1, March 2003.