

시맨틱 웹과 MPEG-7 메타데이터를 이용한 비디오 검색 시스템

이성우, 한종형, 안철범, 나연묵
단국대학교 전자컴퓨터공학과

e-mail : {swlee, jhhan, ahn555, ymnah}@dmlab.dankook.ac.kr

Video Search System using Semantic Web and MPEG-7 Metadata

Seong-Woo Lee, Jong-Hyeong Han, Chul-Bum Ahn, Yunmook Nah
Dept of Electronics and Computer Engineering, Dankook University

요 약

본 논문에서는 시맨틱 웹을 위한 마크업 언어인 RDF와 MPEG-7 표준을 따르는 XML 문서를 이용하여 구축된 비디오 검색 시스템에 대한 연구결과를 기술한다. 본 논문에서 제안된 시스템은 RDF를 통해 비디오 파일에 대한 정보를 웹 자원의 관점으로 정의/저장한다. 또한 비디오에 대한 내용기반 검색을 위해 비디오를 구성하는 중요 프레임들을 추출하여 그림 파일로 저장하는 한편 각 프레임에 대한 정보를 MPEG-7 표준에 따라 XML 형태로 저장한다. 이렇게 구축된 정보들을 바탕으로 비디오 검색 엔진은 사용자의 질의요청을 수행한다. 본 논문에서 제안하는 비디오 검색 시스템은 비디오에 대한 다양한 검색은 물론 비디오에 대한 내용기반 검색이 가능한 장점을 가진다.

1. 서론

인터넷의 인프라 확대 및 멀티미디어 관련 기술의 발전에 따라 VOD(Video On Demand)와 같이 웹을 통한 멀티미디어 서비스의 수요가 급증하게 되었다. 이에 따라 사용자가 원하는 멀티미디어(그중에서도 비디오)에 대한 정확한 검색을 지원하기 위해 비디오에 대한 내용기반(Content-based) 검색 연구가 꾸준히 이루어지고 있다. 이러한 연구에 대한 결과 및 표준으로서 MPEG-7(Moving Picture Experts Group layer-7) 표준안이 제정되면서 비디오에 대한 내용기반 검색을 위한 정보의 표현 및 저장 방법이 통일되었다. MPEG-7 표준은 비디오를 구성하는 다수의 장면(프레임)에 대한 메타데이터를 표현하고 저장하는 지침 일뿐 이러한 프레임들에 대한 특징정보(색상, 모양, 질감)를 추출하거나 구축된 MPEG-7 기반의 정보를 검색하는 것은 표준 외의 것으로 지정해 두고 있다[1]. 그러기에 본 논문에서는 MPEG-7 표준을 기반으로 하여 다른 기술을 접목한 비디오 검색 시스템을 제안하고자 한다.

*본 논문에서는 비디오에 대한 두 가지 관점을 가지고 각각에 대한 비디오 메타데이터 기술 방법을 달리 하고 있다. 첫 번째 관점은 비디오를 하나의 웹 자원(Web Resource)으로서 보는 것이고 두 번째 관점은 비디오를 프레임이라 불리우는 다수의 정지 화상으로 보는 것이다. 이런 두 가지 관점 하에서의 비디오 메타데이터는 각각의 형식으로 표현될 수 있다. 전자는 시맨틱 웹(Semantic Web)을 위한 마크업 언어(Markup Language)인 RDF(Resource Description Framework)[2]를 이용하여 표현되며, 후자는 앞서 언급한 MPEG-7 표준을 따르는 XML(eXtensible Markup Language) 문서를

이용하여 표현된다. 이를 위해 본 논문의 비디오 검색 시스템에서는 비디오 파일에 대한 RDF 및 MPEG-7 표준의 XML 형식의 메타데이터를 생성하여 저장하고 이 데이터들을 이용하여 사용자가 원하는 비디오 또는 비디오의 특정 장면을 검색할 수 있도록 하였다.

본 비디오 검색 시스템은 시맨틱 웹 기술과 MPEG-7 표준을 혼합함으로써 비디오에 대한 검색의 다양성을 제공하는 한편 내용기반 검색을 가능하게 한다. 이를 위해 일반 사용자에게는 웹 페이지 형태의 검색 인터페이스를 제공하며 웹 서비스 인터페이스를 통해 외부 시스템의 검색 요청에 대한 결과를 반환할 수 있도록 하였다. 이에 대한 구체적인 예제로서 본 논문에서는 다음의 질의 "Find scenes of 'The Lord of The Rings 3' that show 'Aragorn' in battle."을 이용하여 질의 입력 방법, 검색 결과 및 메타데이터 생성 과정을 기술한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 본 연구의 근간이 되는 시맨틱 웹과 RDF 및 MPEG-7 표준에 대해 살펴보고 3장에서는 이를 바탕으로 비디오 메타데이터를 구성하는 방법을 기술한다. 4장에서는 전체 시스템의 구조 및 비디오 검색 시나리오 및 검색 결과를 통한 구현을 보이고 마지막으로 5장에서는 결론 및 향후 연구에 대해 기술한다.

2. 관련 연구

2.1 시맨틱 웹과 RDF

* 본 연구는 한국과학재단 목적기초연구(R01-2003-000-10133-0) 지원으로 수행되었음.

Tim Berners Lee에 의해 처음으로 제안된 시맨틱 웹[3]은 메타데이터의 개념을 통하여 웹 자원에 의미 정보를 덧붙이고 이를 이용

하여 컴퓨터 에이전트가 이 의미 정보를 자동으로 추출할 수 있는 패러다임을 조성하는 웹(Web)의 지능적 확장이다. 이를 이용함으로써 의미 정보의 자동 추출뿐 아니라 정보의 확장이나 공유 등도 가능하게 된다. 시맨틱 웹은 RDF라는 XML을 언어적 기반으로 하는 표현 수단을 사용한다. RDF에서는 RDF Triple과 RDF Schema를 통해 자원 간에 의미가 부여되며 이를 사용함으로써 보다 정확한 검색 성능을 제공할 수 있고 지식 공유와 교환이 가능하게 된다 [2][4]. 이렇게 구축되는 정보는 온톨로지(Ontology)라 불리우며 RDF 외에도 OWL(Web Ontology Language)로 표현 가능하다.

RDF는 컴퓨터 에이전트가 이해할 수 있는 웹 메타데이터의 형식을 표현할 뿐이며 개발자는 별도의 어휘(Vocabulary)를 정의하여 온톨로지를 구축해야 한다. 이런 점에서 웹 자원 기술을 위한 메타데이터 표준인 DC(Dublin Core)[5]는 그러한 어휘들을 제공하는 좋은 예이며, [6]에서는 DC의 어휘가 어떻게 RDF 형식으로 표현될 수 있는지를 보여준다.

2.2 멀티미디어 내용기반 정보검색과 MPEG-7 표준

내용기반 하에 다양한 멀티미디어 정보 검색을 효율적으로 지원하기 위한 기술을 개발하고 이를 국제 표준화하고자 하는 것이 MPEG-7이다[7]. MPEG-7은 비디오 데이터 그 자체가 아닌 비디오가 가진 오디오(Audio), 비주얼(Visual) 데이터의 내용에 대한 표현 방법을 다루는 것이다. 이를 다른 말로 ‘메타데이터(Metadata)’, 또는 ‘Bits about bits’라고 표현하기도 한다. 이러한 표현을 위해 MPEG-7에서는 비디오의 특성을 고려하여 비디오를 구성하는 프레임의 기본 단위로 보며, 비디오에서 장면의 전환이 이루어지는 부분인 컷(cut), 컷으로 구분되며 하나의 카메라 동작에 의해 촬영된 작은 비디오 단위인 샷(shot), 논리적인 내용이 같은 연속된 샷으로 이루어진 단위인 에피소드(episode)로 분할하여 각각에 대하여 그 특징이나 의미를 XML 형식으로 수식한다. MPEG-7에서는 주로 오디오, 비주얼 정보(정지화상, 픽처, 그래픽, 3D 모델, 오디오, 스피치, 비디오)의 표현을 그 대상으로 하고 있으나, 이에 대한 특징추출(Feature Extraction)이나 검색에 대한 것은 표준에서 제외된다.

멀티미디어 데이터가 갖는 풍부한 정보에 비추어 보아, MPEG-7에는 두 가지 방향의 접근 및 해법이 제시되고 있다. 그 하나는 상위 레벨 내용[8][10]에 기반한 검색과 다른 하나는 그보다 하위 레벨 내용[9][11]에 기반한 검색 방법이다. 하위 레벨의 내용에는 키워드, 형태, 색, 크기, 위치, 방향성, 멜로디 등이 포함될 수 있고, 상위 레벨의 내용에는 의미 정보들이 포함될 수 있다. 하위 레벨의 내용에 의한 검색은 사용자 질의가 부자연스러운 점은 있으나 시스템이 자동으로 검색할 수 있다는 장점이 있는 반면, 상위 레벨의 내용에 의한 검색은 좀더 자연스런 질의를 할 수 있으나 현재의 컴퓨터 기술로는 사용자의 개입을 필요로 한다는 단점이 있다.

3. 비디오 검색을 위한 메타데이터의 설계 및 생성

3.1 동기 및 설계 범위

본 비디오 검색 시스템의 실현을 위해서는 서두에 언급한 두 비디오 메타데이터가 구축되어야 한다. 그 중 첫 번째를 비디오 온톨로지(Video Ontology)라 부르며 이를 RDF를 통해 표현한다. 두 번째는 MPEG-7 표준을 반영한 XML로 간단히 MPEG7 파일이라 칭한다. 이 두 데이터가 어떻게 구성되느냐에 따라 시스템에 영향을 끼치기 때문에 검색에 이용될 두 데이터에 대한 설계 원칙 및 범위를 결정하였다.

를 결정하였다.

- 첫째, 비디오 온톨로지는 웹 자원의 관점으로 기술되어야 하며 이를 위한 적절한 어휘(vocabulary)를 정의하여 구성한다.
- 둘째, MPEG7 파일은 비디오의 비주얼(Visual)에 대한 상위 레벨의 내용을 담고 있으며 그 기본 단위는 샷(shot)으로 정한다.

이렇게 함으로써 시스템의 기능을 명확히 할 수 있다. 특히 MPEG-7 표준의 방대하고 복잡한 세부 사항들을 모두 만족시키기란 매우 어려운 일이므로 그중 직관적인 이해가 가능하고 비교적 구현이 단순한 사항을 선택하였다. 그러나 향후 연구에서는 본 연구가 다루지 않은 MPEG-7 표준의 다른 내용(Visual에 대한 하위 레벨의 내용)도 추가할 것이다.

3.2 비디오 온톨로지의 설계 및 생성

앞서 언급한 것처럼 비디오 온톨로지의 생성에 앞서 이를 위한 어휘의 정의가 필요하다. 이를 위해서 가칭 VMDH(Visual Media Descriptor for HERMES)라는 시각 미디어의 메타데이터 표현을 위한 요소 집합을 정의하였다(HERMES[13]는 본 연구에서 구현하고 있는 분산 시각 미디어 검색 시스템의 이름이다). VMDH의 현재 버전은 0.1로 이는 웹 자원 기술을 위한 메타데이터 표준인 Dublin Core[5] 및 시각 자원의 기술을 위한 메타데이터 표준인 VRA Core[12]의 요소를 반영하였고 이에 더하여 비디오의 특성을 반영한 요소를 추가하였다. <표 1>에서 VMDH v0.1의 요소들을 보이고 있다.

표 1. VMDH v0.1 요소들

요소	설명
Media URL	비디오 자원(MPEG 1,2,4)의 위치 또는 파일 경로를 나타냄.
MP7 File URL	MPEG-7 XML 파일의 위치를 나타냄.
Title	비디오 자원의 제목(원제목, 부제목, 시리즈명)을 나타냄.
Subject	비디오 자원이 담고 있는 주제, 키워드 또는 카테고리들을 나타냄.
Description	비디오 자원에 대한 설명을 담고 있음.
Creator	해당 비디오의 제작자(감독, 제작사 등)를 표시함. 주로 사람의 이름, 소속, 역할을 기술함.
Contributor	해당 비디오의 배포자(배급사, 소유권자 등)를 표시함. 이 역시 사람이 대상이 됨.
Date	비디오가 제작되거나 공개된 날짜.
Format	비디오의 MIME type, 해상도, 재생시간 등을 정보를 표현함.
Key Frame Images	비디오에서 추출한 여러 shot 중 대표가 되는 key frame 그림 파일들의 위치.
Relation	해당 비디오와 연관된 다른 비디오 자원의 URL 또는 제목
Location	비디오 내에 등장하는 지리적 행정적 위치

이렇게 정의된 요소들은 비디오 온톨로지의 어휘들로 사용되어 웹 자원으로서의 비디오에 대한 수식에 이용된다. 그림 1은 이렇게 구축된 비디오 온톨로지의 예제를 보여준다. 그림에서는 하나의 비디오 자원에 대해 VMDH의 요소를 적용하여 구성된 RDF의 그래프를 보이고 있다. 그래프에서는 “The Lord of The Rings 3, Return

of The King”에 대한 비디오 온톨로지의 내용을 보이고 있고 이를 통해 단순한 비디오 자원에 의미 정보를 덧붙일 수 있다.

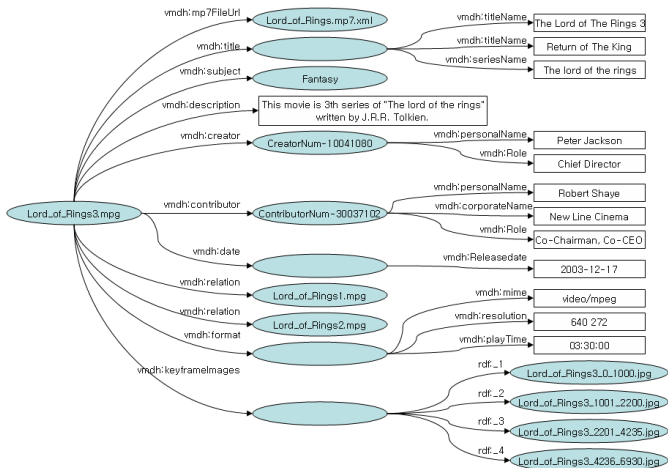


그림 1. 비디오 온톨로지의 그래프 표현

3.3 MPEG-7 메타데이터의 생성

내용기반 검색 지원을 위한 MPEG7 파일을 구축하기 위해 원본 비디오에 대한 샷 단위의 화면분할이 필요하다. 또한 이렇게 분할된 화면 별로 개발자가 의미 정보를 삽입하는 과정이 필요하며 모든 작업 후에는 수식된 결과를 MPEG-7 표준을 따라 XML 형태로 표현 저장한다. 또한 각 샷 단위의 화면 별로 대표 프레임(key frame)을 추출하여 별도로 저장한다. 이를 위해 본 시스템에서는 MPEG-7을 지원하는 도구(Annotation/extraction tool)를 이용하였다. 그림 2는 화면의 샷 단위 분할과 대표 프레임들에 대해 어떠한 의미 정보를 부여하는 지를 보여주고 있다.

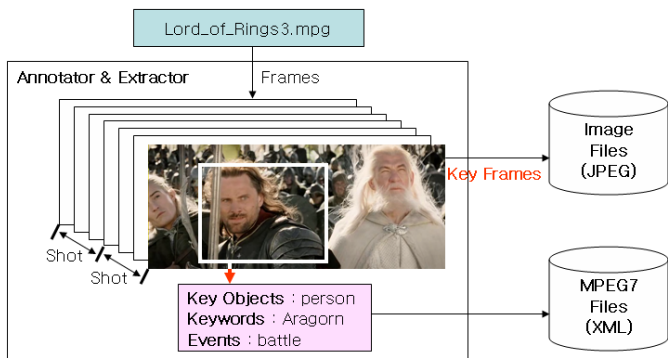


그림 2. 대표 프레임 추출 및 의미 정보의 생성

그림에서는 영화 “The Lord of The Rings 3”의 비디오 파일인 “Lord_of_Ring3.mpg”를 불러들여 의미 정보를 삽입하는 예를 보이고 있다. 샷 단위로 분할된 주요장면들(key frames)에는 상위 레벨의 내용에 해당하는 의미 정보를 삽입하는데 이러한 정보들에 다음을 포함하도록 하였다.

- Key Objects** : 장면에 나오는 개체의 종류
- Static Scene** : 장면의 주요 배경
- Events** : 장면 속에서 일어나고 있는 사건, 행위 또는 자연현상
- Keywords** : Key Objects, Static Scene, Events를 구체적으로 수식하거나 명시할 수 있는 단어

각 주요장면 별로 이러한 정보들을 담아 XML 형식으로 저장하는 한편 해당 그림 파일을 추출해 별도로 저장했다. 이렇게 함으로써 비디오에 대한 상위 레벨의 내용들을 검색할 수 있는 사전 데이터 구축 작업을 수행하였다.

4. 시스템 설계 및 구현

4.1 시스템 구조

앞에서 기술한 것처럼 두개의 비디오 메타데이터를 구성한 후에 이를 해석하여 사용자 질의에 적합한 비디오 및 특정 장면의 정보들 검색하는 시스템을 구축하였다. 그림 3은 본 논문에서 제안하고 구축한 비디오 검색 시스템의 구조를 보여준다.

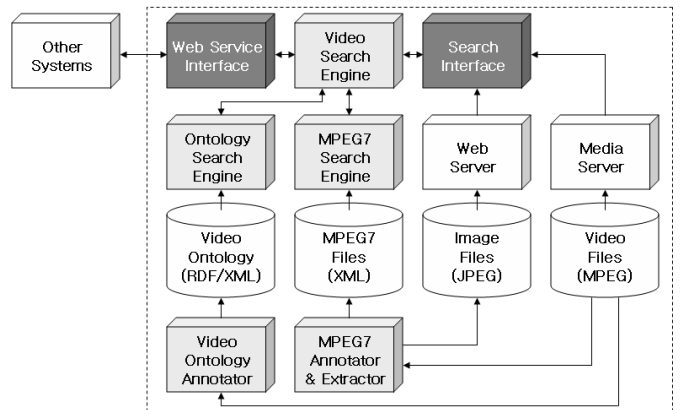


그림 3. 비디오 검색 시스템의 구조

본 시스템은 검색 및 비디오 서비스를 위한 4개의 데이터 저장소와 2개의 비디오 메타데이터 생성 도구, 그리고 비디오 검색을 위한 3개의 검색 엔진과 2개의 인터페이스로 구성되어 있다.

- Video Search Engine** : Search Interface 또는 Web Service Interface로부터 질의를 받아 검색된 비디오 정보를 반환한다.
- Ontology Search Engine** : Video Search Engine으로부터 비디오 자원에 대한 질의가 올 경우 Video Ontology를 검색하여 해당 비디오 정보를 반환한다.
- MPEG7 Search Engine** : Video Search Engine으로부터 비디오의 특정 장면에 대한 질의가 올 경우 MPEG7 파일을 검색하여 해당 비디오 및 장면에 대한 정보를 반환한다.
- Video Ontology** : 3.2절에서 설명한 비디오 온톨로지의 저장소로 RDF/XML 파일로 이루어져 있다.
- MPEG7 Files** : 3.3절에서 설명한 MPEG-7 메타데이터의 저장소로 XML 파일로 이루어져 있다.
- Image Files** : 비디오에서 추출한 주요장면들(key frames)을 JPEG 파일 형태로 모아둔 저장소이다.
- Video Files** : 검색될 비디오 원본 파일들로 Media Server를 통해 사용자에게 제공되며 Video Ontology Annotator와 MPEG7 Annotator & Extractor를 통해 검색을 위한 사전 데이터를 추출할 수 있는 MPEG 파일 형태의 저장소이다.
- Video Ontology Annotator** : 비디오 온톨로지를 RDF/XML 형태로 작성해 주는 편집도구이다.
- MPEG7 Annotator & Extractor** : 비디오로부터 주요장면에 대한 의미정보를 편집해 주며 장면에 대한 이미지 파일을 추출하는 편

집도구이다.

·**Search Interface** : 웹 페이지 형태의 사용자 인터페이스이다.

·**Web Service Interface** : 다른 시스템에 본 검색 서비스를 제공하기 위한 웹 서비스 형태의 인터페이스이다.

4.2 검색 시나리오 및 결과

예제 질의 "Find scenes of 'The Lord of The Rings 3' that show 'Aragorn' in battle."의 경우, 본 시스템에서는 자연어의 직접 입력은 지원되지 않으므로 이를 검색 인터페이스의 검색 필드에 맞게 다음과 같이 분할하여 질의를 생성해야 한다.

·**Title** : The Lord of The Rings 3

·**Key Objects** : person

·**Event** : battle

·**Keywords** : Aragorn

위와 같이 표현된 질의는 웹 페이지 형태의 Search Interface에 입력되며 이에 대한 검색을 수행하면 그림 4 와 같은 결과를 보이게 된다.

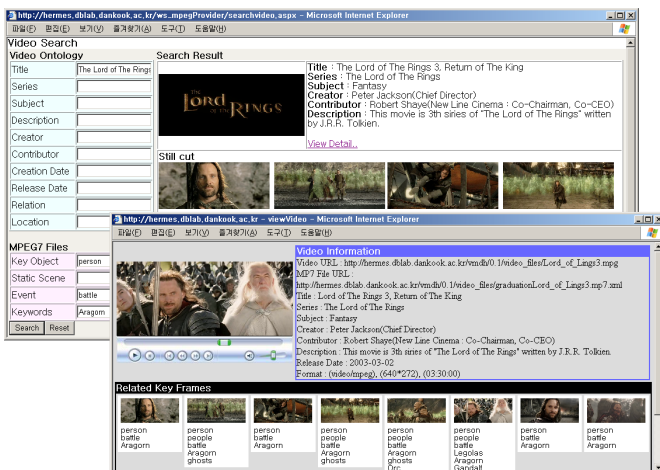


그림 4. 검색 인터페이스를 통한 검색 결과

그림의 상단은 사용자의 질의를 입력받아 검색된 비디오 목록을 보여준다. 해당 비디오에 대한 보다 구체적인 정보를 얻거나 비디오를 관람하기 위해 비디오 목록을 선택하면 그림 하단에서 보이는 것처럼 비디오에 대한 세부적인 내용 및 사용자 질의에 부합하는 주요장면을 보여주며 해당 장면으로 바로 접근할 수 있도록 해준다.

5. 결론 및 향후 연구

MPEG-7은 멀티미디어 내용기반 정보 검색을 지원하기 위한 데이터의 표현에 대한 표준으로 이에 대한 정보 추출 및 검색 방법에 대해서는 관여하지 않는다. 본 비디오 검색 시스템은 MPEG-7 표준을 이용함은 물론 시맨틱 웹의 개념을 적용한 색다른 비디오 검색 기법을 이용하여 비디오에 대한 내용기반 검색을 지원하는 한편 보다 폭넓은 검색을 가능하게 한다. 본 논문에서는 이를 위한 비디오 메타데이터 구축 및 검색 시스템을 제안하였고 이를 통한 비디오 검색 예를 보였다.

그러나 본 비디오 검색 시스템은 비디오의 하위 레벨의 내용에 대한 검색을 지원하지 못 하고 있으며, 구축된 비디오 온톨로지를 시맨틱 웹의 진정한 목적인 자동적인 의미정보 추출을 위해 사용하지 못 하고 있다. 향후 연구에서는 이를 개선하여 비디오의 특징 정보인 색상, 모양, 질감 등의 하위 레벨의 내용 검색에 대해 접근할 것이며 비디오 온톨로지의 개선을 통해 자동적인 비디오 관련 정보 추출이 가능하도록 노력할 것이다.

참고문헌

- [1] Jose M. Martinez, "Overview of the MPEG-7 Standard", Doc. ISO/IEC JTC1/SC29/WG11 N3158, December 1999, Maui, Hawaii
- [2] <http://www.w3.org/TR/rdf-primer/>
- [3] Berners-Lee, T., Hendler, J. and Lassila, O., "The Semantic Web," Scientific American, 2001.
- [4] Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M. and Horrocks, I., "The Semantic Web: the roles of XML and RDF," IEEE Internet Computing, Vol. 4, No. 5, pp.63-73, 2000.
- [5] Dublin Core Organization, Dublin Core Metadata Initiative, <http://dublincore.org/index.shtml/>.
- [6] Dave Beckett, Eric Miller, Dan Brickley, "Expressing Simple Dublin Core in RDF/XML", <http://dublincore.org/documents/dcmes-xml/>
- [7] "MPEG-7: Context and Objectives", ISO/IEC JTC1/SC29/WG11 N2460, October 1998, Atlantic City, USA
- [8] "Introduction to High-level Visual Description Tools", Doc. ISO/IEC JTC1/SC29/WG11 MPEG2005/N7320, July 2005, Poznan, Poland
- [9] "Introduction to Low-level Visual Description Tools", Doc. ISO/IEC JTC1/SC29/WG11 MPEG2005/N7319, July 2005, Poznan, Poland
- [10] "Introduction into MPEG-7 Audio Tools", ISO/IEC JTC1/SC29/WG11/ N7709, October 2005, Nice, France
- [11] "Description of MPEG-7 Audio Low Level Descriptors", ISO/IEC JTC1/SC29/WG11/ N7709, October 2005, Nice, France
- [12] VRA Core Version 3.0, <http://www.vraweb.org/vracore3.htm>
- [13] Yunmook Nah, Bogju Lee, Jungsun Kim, "Visual Media Retrieval Framework using Web Service," LNCS 3597, Springer Verlag 2005, pp.104-113. (Proc. HSI 2005, July 2005, Tokyo, Japan).