



KBS 비디오아카이브 시스템과 메타데이터

+ 임태현 KBS 방송시설국 제작시설부



소재를 찾기 위해 인터넷에서 검색을 통해 영상의 제목을 보고 한참 동안 파일을 다운로드 받았는데, 막상 해당 영상 파일을 재생해보니 내용이 제목과 달랐던, 유쾌하지 않은 경험을 해본 사람이 적지 않을 것으로 생각한다. 일반적으로 영상 콘텐츠는 에센스와 메타데이터로 구성되는데, 위의 경우는 이 에센스와 메타데이터가 일치하지 않아서 발생한 사례라고 볼 수 있다. 더구나 영상은 저장 매체가 파일로만 국한되지 않고 전통적인 필름이나 자기테이프 등의 다양한 형태로 존재해왔고, 동일한 내용을 담고 있는 영상이라 하더라도 쓰임새에 따라 방송본, 촬영본, 편집본과 같은 다양한 구분이 존재한다. 굳이 매체와 형태의 다양성이 아니더라도, 지금은 SNS 등을 통해 사용자와 제작자의 구분 없이 수많은 미디어가 제작되고 소비되는 빅데이터의 세상이다.

산더미처럼 쌓여가는 미디어 속에서 알맞은 영상을 적시적기에 찾아내고 활용하기 위해서는 메타데이터에 대한 적절한 관리가 반드시 필요하다. 저장해 놓은 영상을 찾아낼 수 없고, 재생되는 영상이 어떤 영상인지 알 수 없다면 그

건 그냥 무의미한 이진수 데이터의 집합체일 뿐이다. KBS에서는 방송영상의 효율적인 활용을 위한 메타데이터의 체계적인 관리와 영상의 안전한 보존을 위해 비디오아카이브 시스템을 구축해서 활용하고 있다. 이 글에서는 KBS 비디오아카이브 시스템에서 메타데이터가 어떻게 만들어지고 효율적인 활용을 위해 어떤 과정을 거치게 되는 지에 대해서 기술해보고자 한다.

메타데이터란?

위키피디아(<http://ko.wikipedia.org/wiki/메타데이터>)에서는, 메타데이터의 정의를 어떤 구조화된 데이터를 분석, 분류하고 부가적 정보를 추가하기 위해 뒤에 함께 따라가는 데이터라고 내리고 있다. 이를테면, 사진이라는 이미지 데이터를 찍어 기록할 때 함께 기록하는 카메라 기기 정보, 촬영 시간, 사진 크기, 사진 촬영지 GPS 정보 등과 같은 부가적인 성격의 데이터를 메타데이터의 예로 들 수 있다. 이런 메타데이터는 콘텐츠 파일 안에 에센스와 함께 기록될 수도 있고, 별도의 파일로 존재할 수도 있다.

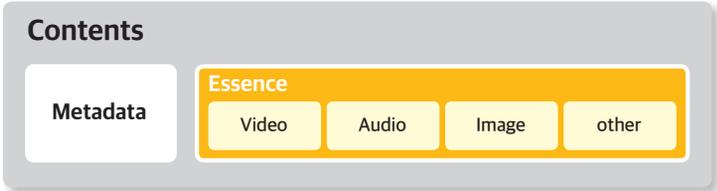


그림 1. 콘텐츠, 메타데이터 그리고 에센스의 관계

방송영상 메타데이터란?

그렇다면, 방송영상 메타데이터란 무엇인가? 그 전에 일단, 방송용 영상 콘텐츠와 일반 미디어 영상 콘텐츠의 큰 차이점을 짚어보자. 방송용 영상 콘텐츠는 일반 미디어 영상 콘텐츠와 비교할 수 없을 정도로 큰 대용량 파일이다. 대략 50Mbps급의 1시간짜리 방송 영상의 파일크기가 약 25GB에 이를 정도다. 더구나 하루에도 수십 개의 프로그램을 제작하고 방송을 하는 방송사의 특성상, 이런 대용량 파일들이 매일 수십 개가 생성되고 보존해야 할 대상이 된다. 이런 환경 속에서 방송메타데이터의 역할은 소비자의 입장에서 보존된 대용량 데이터 안에서 원하는 대상 데이터 검색을 제대로 할 수 있도록 하고 관리자의 입장에서는 보유 중인 데이터의 분류, 저장, 관리, 제어를 잘할 수 있도록 하는 것이다.

방송과 관련된 메타데이터 분야의 표준으로는 MPEG-7, TV Anytime 표준, SMPTE 표준 등이 있다. MPEG-7은 오디오 및 비디오(AV, Audio-Visual) 콘텐츠에 대한 일반적인 메타데이터 표준으로 MPEG-1, 2, 4 등의 동영상 압축 부호화 표준을 제정한 MPEG(Moving Pictures Expert Group)에서 마련했다. 원래 취지는 인터넷상의 MPEG 표준에 의한 압축 AV 콘텐츠를 쉽게 검색할 수 있도록 다양한 서술 정보의 틀을 표준화하는 것이다. TV Anytime 표준은 PVR(Personalized Video Recorder)을 위한 메타데이터 표준으로써, 큰 저장장치(하드 디스크)를 갖는 AV 단말기에 저장된 콘텐츠를 쉽게 검색할 수 있게 하는 데 목적이 있다. 이후 TV Anytime 표준은 방대한 저장 장치를 갖는 AV 서버들로부터 통신망을 통하여 원하는 콘텐츠를 검색한 다음 전송받아 재생하는 네트워크 연결형 PVR(Networked PVR)을 위한 메타데이터까지도 포함하게 됐다. SMPTE는 미국의 영화 및 텔레비전 공학회로서 주로 영화 및 스튜디오 내 TV 콘텐츠에 대한 메타데이터를 표준화했다. 이러한 메타데이터들은 범용적으로 응용될 수 있는 형태로 표준화된 것으로 실제 지상파, 위성, 케이블 TV, 지상파-위성 DMB 등의 방송 서비스별로 이러한 범용 표준으로부터 상당 부분을 차용하고 각 서비스에 특화된 일부 메타데이터를 추가로 정의하여 별도의 표준들을 마련해 사용하고 있다.

KBS 방송 메타데이터 표준

메타데이터에는 보통 타이틀, 내용, 길이, 크기, 제작진, 저작권, 방송일, 촬영일, 제작일, 매체, 식별기호 등과 같이 다양한 정보가 미디어 제작 회사별로, 시스템별로 제각각 존재한다. 더구나 위에서 언급한 국제 표준들에서는 여러 가지 메타데이터가 정의되어 있으나, 방송사가 이를 모두 생성하려면 제작비가 과다하게 소요되므로, 꼭 필요한 수준으로 서비스되는 것이 현실적이다. 이에 따라 세계 유수의 공영방송사들은 자사에 맞는 전사적인 메타데이터 체계를 수립하여 활용하고 있다. 영국의 공영방송사인 BBC는 관련 시스템 간의

데이터 공통성 증진을 위해 표준 메타데이터 교환프레임워크의 데이터 모델인 SMEF-DM(Standard Media Exchange Framework Data Model)을 정의하여 방송사 내 메타데이터 공유 체계를 마련하였고, 일본 공영방송사인 NHK는 국제 메타데이터 표준인 MPEG-7을 기반으로 데이터 모델을 정의하고 이를 바탕으로 MPF(Metadata Production Framework)라고 하는 메타데이터 제작시스템의 구성 요소와 메타데이터의 형식, 조작 인터페이스를 규정하는 공통 기반 체계를 마련하고 있다. 이와 같은 노력과 마찬가지로 KBS에서도 일관성 있는 메타데이터 공유를 위해 사내 제작 시스템, 보도 시스템, 편성/송출 시스템, 아카이브 시스템을 아우르는 전사적인 메타데이터 체계를 수립하고 각 사내 시스템들은 서로 표준화된 메타데이터를 통해 호환성 있게 공유하고 활용되고 있다.

KBS 비디오아카이브 시스템 소개

KBS에서는, 기존 테이프를 보유하고 있는 영상 콘텐츠들에 대한 디지털화 작업과 효율적인 콘텐츠 관리를 위해 메타데이터 표준 정립을 하고 방송 콘텐츠 보존성 및 재활용성을 향상하기 위해 지난 2011년에 비디오아카이브 시스템을 개발, 오픈하였다.



그림 2. KBS비디오아카이브 사용자 웹 화면

사용자는 아카이브 시스템의 웹 페이지에 접근해서 필요한 영상을 검색하여 원하는 부분만 부분인출하거나 전체 콘텐츠를 연계 시스템으로 다운로드를 받을 수 있고, 테이프를 대출받을 수도 있다. 이때 사용자는 단순히 원하는 검색어 입력만으로 검색을 할 수 있지만, 상세검색을 통해 보다 세밀한 조건으로 검색할 수도 있다. 상세검색에서는 방송일/제목/인물/내용을 지정한 후 검색어와 조합하여 검색을 하는 상세 옵션 검색, 방송일과 채널을 선택하여 원하는 프로그램을 검색하는 편성표 검색, EBU의 ESCORT 2007(EBU System of Classification Of Radio and Television programmes)을 참고하여 KBS 상

항에 맞게 체계화한 표준분류 값을 통한 검색, 뉴스/다큐멘터리/교양/문화/드라마/연예오락/영화만화/스포츠 등의 아카이브 장르를 통한 검색, 편성별칭을 통한 검색, 주제어 사전의 주제어를 사용한 검색, 영상 콘텐츠 개별 미디어들에 부여된 고유한 아카이브 ID를 사용한 검색, 미속촬영/초고속촬영/수중촬영/근접촬영 등의 특수촬영영상 검색과 같은 다양한 검색 방법을 제공한다.



그림 3. KBS 비디오아카이브 사용자 업무 흐름도

비디오아카이브 KDAS는 KBS 콘텐츠의 최종 저장/보관 시스템이며 콘텐츠 재활용을 효율적으로 지원하기 위한 시스템으로 사내의 많은 시스템들과 콘텐츠를 주고받는 연계를 하고 있으며, 연계 대상 시스템들은 다음의 그림과 같다.

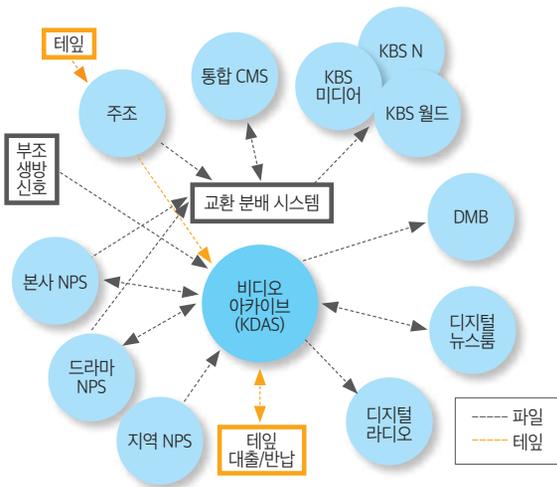


그림 4. KBS 비디오아카이브 시스템 연계도

메타데이터 입출력 흐름

앞서 언급한 KBS 비디오아카이브 시스템에서 사용자가 콘텐츠를 검색할 수 있도록 제공되는 다양한 검색 방법은 KDAS 시스템에서 관리되고 있는 메타데이터를 기반으로 한다. KBS에서는 아카이브 관리부에서 콘텐츠 서비스 운

영과 저작권/사용제한, 주제어 등록, 출연자/제작자/스태프 관리, 세그먼트 작업, 참조자료 등의 메타데이터 관리에 대한 역할을 담당하고 있다. 따라서 비디오아카이브 시스템에서 사용자가 검색한 콘텐츠에 대한 상세 화면을 보여 줄 때, 아카이브 관리부에서 직접 입력/관리하는 메타데이터가 사내 메타데이터 허브 격인 KBS 디지털 편성정보 시스템에서 제공하는 Open API로 조회한 메타데이터와 함께 사용되고 있다. 또한 비디오아카이브 시스템에서 관리하는 메타데이터는 사내 타 시스템과 연계를 할 때 에센스 파일과 함께 별도의 XML 파일로 전달된다.



그림 5. KBS 비디오아카이브 사용자 상세검색 화면

메타데이터 전송 효율을 위한 노력

KBS 비디오아카이브 시스템의 메타데이터는 내부의 RDBMS에 저장되어 관리되고 있다. 이렇게 관리되는 메타데이터가 타 시스템과 연계될 때는 XML 파일을 사내 표준 인터페이스인 SOAP 기반의 통신으로 송수신하고 있지만, 비디오아카이브 시스템상에서 사용자 요청을 처리할 때는, 메시지 경량화를 통한 성능 향상을 목적으로 REST 통신을 통해 JSON 경량 메시지를 송수신하고 있다.

```

http://localhost:8080/Json/SyncReply/Contacts
{
  "Contacts": [
    {
      "FirstName": "Denise",
      "LastName": "Belliot",
      "Email": "denise.belliot@gmail.com"
    },
    {
      "FirstName": "Steve",
      "LastName": "Jobs",
      "Email": "steve@apple.com"
    },
    {
      "FirstName": "Steve",
      "LastName": "Ballmer",
      "Email": "steve@microsoft.com"
    },
    {
      "FirstName": "Eric",
      "LastName": "Schmidt",
      "Email": "eric@google.com"
    },
    {
      "FirstName": "Larry",
      "LastName": "Ellison",
      "Email": "larry@oracle.com"
    }
  ]
}

http://localhost:8080/Xml/SyncReply/Contacts
<ContactsResponse xmlns:i="http://www.w3.org/2003/11/Serialization" xmlns="http://schemas.datacontract.org/2004/07/System.Collections.Generic.IEnumerable`1[[System.Collections.Generic.Contact]]">
  <Contacts>
    <Contact>
      <Email>denise.belliot@gmail.com/</Email>
      <FirstName>Denise/</FirstName>
      <LastName>Belliot/</LastName>
    </Contact>
    <Contact>
      <Email>steve@apple.com/</Email>
      <FirstName>Steve/</FirstName>
      <LastName>Jobs/</LastName>
    </Contact>
    <Contact>
      <Email>steve@microsoft.com/</Email>
      <FirstName>Steve/</FirstName>
      <LastName>Ballmer/</LastName>
    </Contact>
    <Contact>
      <Email>eric@google.com/</Email>
      <FirstName>Eric/</FirstName>
      <LastName>Schmidt/</LastName>
    </Contact>
    <Contact>
      <Email>larry@oracle.com/</Email>
      <FirstName>Larry/</FirstName>
      <LastName>Ellison/</LastName>
    </Contact>
  </Contacts>
</ContactsResponse>
    
```

그림 6. JSON vs XML - 동일한 메타데이터 크기 비교

또한, 사용자 요청이 빈번한 메타데이터는 캐시를 이용한 부하 분산 및 읽기 성능 향상을 꾀하고 있다. 캐시는 나중에 요청 올 결과를 미리 저장해 두었다가 빠르게 서비스해주는 방법을 말한다. 일반적으로 메모리는 디스크보다 읽기 성능이 100배 이상 빠르다. 메타데이터에 대한 쓰기 작업은 RDBMS에 하고, 읽기 작업을 할 때 Spring AOP를 사용해서 캐시 히트 시에는 캐시에서 조회한 값을 읽고, 캐시 미스 시에는 RDBMS에서 데이터를 읽고 캐시로 올린다. KBS 비디오아카이브 시스템에서는 데이터 유형이나 연계대상에 따라 Ehcache나 Redis와 같은 메모리 캐시 기반의 솔루션을 사용하여 입출력 성능 향상을 꾀하고 있다.

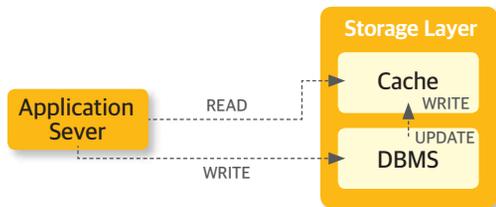


그림 7. 부하 분산 및 읽기 성능 개선을 위한 캐시 사용

추가로, 대용량의 에센스 파일을 송수신할 때는, FTP를 이용한 송수신을 하지만, 메타데이터를 전달할 때나 타 시스템과 SOAP 인터페이스를 이용한 통신을 할 때는 웹서버 단에서 데이터 압축/해제를 하도록 하여 송수신 메시지 사이즈를 줄여 성능향상을 꾀하고 있다.

메타데이터 품질 관리를 위한 노력

KBS 비디오아카이브 시스템에서는 메타데이터 품질 관리를 위해 다음과 같은 노력을 하고 있다. 우선, KBS 사내 타 시스템에서 비디오아카이브 시스템으로 콘텐츠를 이관할 때 반드시 기록을 해서 보내야 하는 메타데이터 항목을 선정하여 이에 미달하는 콘텐츠는 이관을 받지 않는 정책을 취하고 있다. 시

스템적으로는 그림의 관리자 페이지의 메타데이터 품질관리와 같은 기능을 사용하여 프로그램별 대표정보/회별 정보/세그먼트 정보/참조자료와 같은 항목의 점검 및 수정을 할 수 있으며, 대표프로그램 별로 필수로 입력을 해야 하는 메타데이터 대상 항목에 대한 설정은 프로그램 작업규칙 설정 기능을 통해 하고 있다. [그림 8]

또한, 사용자 입장에서 보다 정확한 검색을 할 수 있도록 돕기 위해 사전 관리 기능을 통해 대표어/참조어/키워드에 대한 입력 및 맵핑 작업을 관리를 다음과 같이 하고 있다. [그림 9]

맺음말

지금까지 KBS 비디오아카이브 시스템에서 메타데이터가 어떤 목적으로, 어떠한 과정을 통해 만들어지고 활용되는지를 살펴보았다. 이 글에서 살펴본 내용 이외에도, 현재 KBS 비디오아카이브 시스템에서는 관리하고 있는 에센스의 메타데이터 정합성을 보다 높이기 위한 방편으로 데이터를 특정 패턴으로 검색하고 수정할 수 있는 메타데이터 진단기능을 개발 중에 있다. 또한, 빅데이터 시대가 되고 부상한, 비정형 데이터 관리에 강점을 보이는 기술인 NoSQL의 특성을 활용하여 KBS 비디오아카이브 시스템에서 종전보다 유연한 메타데이터 관리가 가능하도록 적용할 수 있는 범위에 대한 연구를 진행하고 있다. 조심스럽게 말해보는 필자의 개인적인 견해로, 향후 대용량 에센스와 메타데이터의 관리는 레거시 공유 저장 장치와 RDBMS의 활용하던 환경에서 분산 파일 시스템 기반의 솔루션과 NoSQL을 활용하는 환경으로 옮겨가지 않을지 예측해보면서 본 글을 마감하고자 한다.

참고자료

위키피디아 / KBS 방송메타데이터표준 / Java 개발자를 위한 XML, 프리렉 / 빅데이터 시대를 위한 NoSQL 핵심가이드, 로드북 / 방송영상 메타데이터의 현재와 미래, DBguide.net



그림 8. KBS 비디오아카이브 메타데이터 품질관리 화면



그림 9. KBS 비디오아카이브 시스템 주제어 관리화면