

[사전 토론문] 공공 영상문화유산 아카이브 정책과 거버넌스
- 기술적 고려 지점에 대하여

강진욱 (주)제머나이소프트 대표

이번 “새로운 공공 영상문화유산 아카이브 정책과 거버넌스” 세미나는 영상을 문화유산으로 인정하고 이를 국가 기록 유산으로써 저장하고 이를 국민들이 자유롭게 검색하는 시스템을 만들기 위한 정책 방향을 제시하는데 그 목적을 가지고 있습니다. 그동안 영상에 대한 아카이브 시스템은 국가기록원, 대통령기록관과 같은 국가(정부)도 시행하고 있지만 주로 영상을 활용하여 이윤을 생산하는 방송사들이 중심으로 해오고 있다고 해도 과언이 아닙니다. 이를 국가 차원에서의 아카이브 시스템으로 구축할 때 현재 국내 방송사의 아카이브 시스템과 비교하여 다르게 고려되어야 할 기술적인 지점과 국가차원의 아카이브 시스템으로 얻을 수 있는 새로운 기술에 대해 첨언을 하고자 합니다.

1. 공공 영상문화유산 아카이브가 지향해야 할 방향성

공공 영상문화유산 아카이브 시스템은 기본적으로 오픈시스템이어야 합니다. 저 역시 솔루션을 개발하고 납품하는 업체에 소속된 사람이지만 본 시스템은 향후에 어떤 업체의 새로운 제품과 기술을 도입하더라도 붙일 수 있는 오픈 플랫폼을 구축해야 합니다. 이를 위해 오픈 시스템과 아키텍처를 기반으로 시스템을 구축하고, 손쉽게 새로운 기능을 본 시스템에 추가할 수 있어야 합니다. EMA(Entertainment Merchant Association)의 Mezzanine 파일 관련한 문서가 대표적인 예가 될 수 있을 것 같습니다. EMA가 지원하는 Mezzanine 파일에 대한 포맷을 공개하고 ffmpeg 등의 오픈소스를 이용하여 만드는 법을 공개함으로써 업체의 이익을 대변하기보다 업체에 상관없이 EMA 시스템을 활용할 수 있도록 하였습니다. 이와 같은 정보의 공개를 통한 오픈시스템을 구축하는 것은 대단히 중요한 방향성일 것입니다.

또한 본 시스템은 미래 지향적(Future Oriented)인 시스템이어야 합니다. 현재 상용화되어 있는 기술 외의 새로운 신기술을 항상 받아들일 수 있는 시스템이어야 하며, 현재의 안정성보다 오히려 미래의 파괴적 혁신(Disruptive Innovation)적인 기술들을 고려하여야 합니다. REST, SOAP 등의 API를 제공하여 다양한 사업 모델을 개발할 수 있어야 하고, 방송사들이 적극적으로 영상을 업로드하고, 다운로드해서 재사용할 수 있는 투명한 확장성을 제공해야 합니다. 페이스북이나 유튜브와 같은 사이트들이 다른 사이트들에 비해 성능이나 화면이 뛰어나서 성공한 것만은 아닐 것입니다. 사용자들이 개발자들이 직접 접근하여 기능을 개발하고 이를 이용한 에코 시스템이 구축되어 전 세계를 아우르는 플랫폼이 되었습니다. 본 아카이브 시스템 역시 한국에서의 영상 아카이브와 검색, 활용을 위한 플랫폼이 되기 위한 API 제공과 표준 프로토콜을 제공하여 플랫폼으로 성장해야 합니다.

그럼에도 불구하고 이 기술은 완결된 하나의 시스템으로써 사용자와 사업자들에게 서비스를 제공하여야 합니다. 만약 이 시스템이 API만 존재하는 개발자용 시스템이거나 일부 사용자

만을 위한 닫힌 시스템이기보다는 완결된 서비스를 국민에게 제공하는 시스템이었을 때 그 존재 의미가 확연해집니다.

마지막으로 본 시스템을 구축함으로써 해외에 자랑하고 지원할 만한 시스템으로 구축하여야 합니다. 예를 들어 프랑스가 미디어 자산 관리 측면에서 많은 솔루션 업체들을 보유한 것은 INA와 같은 정부 정책에 따른 제품들을 개발하고 발전시켜오면서 세계 시장에서 경쟁했기 때문입니다. 이와 같이 본 시스템은 국내 기술 발전을 이끌 수 있어야 하고, 영상문화유산에 대한 아카이브의 모범으로 국위를 선양할 수 있어야 할 것입니다.

2. 공공 영상문화유산 아카이브에 있어서 새로운 아키텍처의 도입

최근 IT 시장에서 데이터 저장/서비스 시스템 구축의 대세는 클라우드(Cloud)라고 할 수 있습니다. 하지만 아마존, 오라클, 마이크로소프트 등의 클라우드에 저장하는 데이터는 용량 및 전송량에 따른 종량제 과금 정책이 기본이고 전송에 인터넷 구간에서의 시간이 소요된다는 약점 때문에, 아카이브 시스템 구축 시에 클라우드를 도입한 국내 사례가 적은 것이 사실입니다.

그럼에도 불구하고 살펴보면, 현재 글로벌 기업들의 클라우드 전략에는 방송을 포함한 영상 아카이브 시스템을 클라우드 공간에 구축시키는 것이 포함되어 있으며, 실제 방송사들을 대상으로 클라우드를 도입하는 것이 비용을 절감할 수 있다고 마케팅하며 시장 진입을 시도하고 있는 것을 확인할 수 있습니다. 또한 외산 아카이브 솔루션들의 솔루션도 기본적으로 클라우드 연동 기능을 제공하며 기술 시장에서의 패러다임 변화를 꾀하고 있는 것 역시 시장 변화의 흐름으로 읽어야 할 것입니다.

현재 방송사와 공공기관들의 전통적 아카이브 시스템이 내부 사용자만을 위한 일종의 오프라인(offline) 아카이브 시스템이라고 하였을 때, 공공 영상문화유산 아카이브 시스템이 구축 되면, 기존의 방송사와 정부기관들의 아카이브와 다르게 인터넷을 통한 검색 및 영상 활용을 위한 서비스를 반드시 제공해야하는 온라인(online) 아카이브로써의 성격을 가질 수 밖에 없습니다. 이를 위해 전통적인 방식의 데이터센터형 아카이브 시스템 구축과 독립적으로 클라우드 공간을 사용하는 것은 필수적으로 보여집니다.

하지만 이렇게 대규모 시스템에 대한 아카이브 전략이 고민된 적 없는 상황에서 클라우드 환경을 무턱대고 도입하는 것은 큰 문제가 있을 수 있습니다. 수십~수백 페타바이트(PB)에 달할 용량, 이에 따르는 비용과 전송에 소요되는 시간(성능)의 문제를 먼저 고민해야 하고 이에 따르는 운영, 프로세스, 보안, 거버넌스에 대한 사전 조사와 테스트가 필수적입니다. 그리고 이 시험 결과를 바탕으로 클라우드 아키텍처를 선택해야 할 것입니다.

공공 영상문화유산 아카이브 시스템은 기존의 방송사와 공공기관들이 구축했던 전통적인 시스템을 퍼블릭 클라우드와 결합하는 실용적 하이브리드 클라우드(pragmatic hybrid cloud)에 대한 국내 최초의 대규모 실증적 구축 사업이 되어야 할 것이며, 이는 국내는 물론 전세계에 모델이 되어야 한다고 봅니다.

3. 아카이브와 메타데이터 입력

현재 해외에서는 딥러닝(Deep Learning) 기술을 이용하여 메타데이터를 자동으로 추출하는 기술에 대한 연구와 상용화가 이루어지고 있습니다. 구글에서 베타테스트를 시작한 Video Intelligence API 기술이 대표적인 예이며, 상용화 솔루션으로는 Veritone과 같은 솔루션들이 있습니다.

2000년대 초반 아카이브에 대한 논의가 최초에 시작되던 당시에 많은 논문은 장면 전환 지점 추출(Scene Change Detection)에 관한 것이었지만 최근에는 영상 분석 기술은 딥러닝 기술을 기본으로 하고 있습니다. 일부 기술들은 이미 인간이 추출하는 메타데이터보다 더 많은 데이터를 추출합니다. 일례로 Veritone과 같은 솔루션은 등장하는 사람에 대한 정보, 입고 있는 옷의 상표, 목소리, 자동차의 번호판 등 다양한 정보를 획득하고 이를 색인화하는 단계에 이르렀습니다.

공공 영상문화유산 아카이브 시스템을 구축할 때에는 아마 이러한 자동화된 메타데이터 입력 기술에 대한 적용 또는 연구 작업도 동반되어야 할 것입니다. 자동화된 메타데이터 입력 기술을 포함하여 딥러닝 기술과 공공 영상문화유산 아카이브 시스템을 함께 그려보았을 때, 우리가 그려보아야 할 방향은 두 가지가 있지 않을까 합니다.

첫 번째는 메타데이터 입력을 콘텐츠 프로바이더(Provider)에 전적으로 기대거나 인력을 투입하지 않더라도 가능한 최대의 데이터를 추출하여 저장할 수 있는 메타데이터 자동화 시스템의 구축일 것입니다. 딥러닝 기반의 메타데이터 추출 시스템은 데이터가 늘어날수록 더욱 많은 메타데이터를 추출할 수 있을 것이기에 구축 시스템과 그 운영을 통해 지속적으로 성능이 향상되는 메타데이터 시스템을 구축할 수 있을 것입니다.

즉 사람이 투입되지 않더라도 점차적으로 메타데이터가 늘어나고 딥러닝을 통해 기술적 완성도가 높아가면서 데이터가 더욱 충실해지는 효과를 얻을 수 있을 것입니다. 부가적으로 시스템 운영 비용에서 메타데이터와 관련한 인건비를 절감할 수 있는 것은 당연한 부가 효과가 아닐까 합니다.

두 번째는 아카이브에 저장된 데이터를 다양한 영상 분석 연구소들에 제공해야 한다는 것입니다. 최근 국내의 모든 영상 분석 연구자들은 유튜브나 네이버 동영상 등의 데이터를 API를 통해서 사용하고 있습니다. 이 API들은 대규모 연구에 사용하기에는 많은 제약점들을 가지고 있습니다. 이러한 정보의 독점은 연구 기관들의 연구 의욕을 떨어뜨리고, 새로운 영상 분석 모델이 나오기에 힘든 환경을 만듭니다. 공공 영상문화유산 아카이브 시스템이 오픈되고, 영상 데이터가 학술적 목적/연구적 목적으로 제공하게 된다면 국내 기술이 해외 기술보다 앞설 수 있는 기반이 되지 않을까 생각합니다.

공공 영상문화유산 아카이브의 방향에 대한 정책적 방향 및 이슈를 고려하면서 함께 새롭게 얻을 수 있고 발전시킬 수 있는 이러한 기술 방향에 대해서 항상 놓치지 않고 진행되었으면 하는 노파심에서 몇 마디를 주제넘게 첨언하였습니다.

감사합니다.