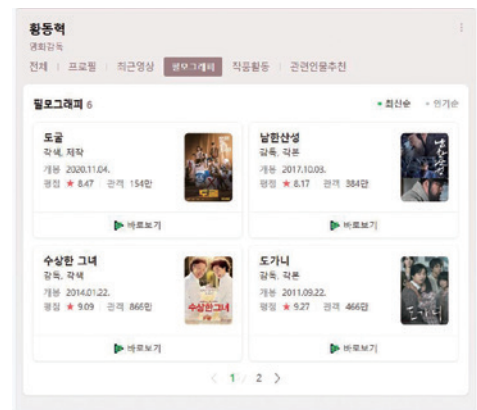


메타데이터의 중요성

글. 이형극 한국전자통신연구원(ETRI) 선임연구원

최근, 제74회 프라임타임 에미상에서 비영어권 작품 최초로, 넷플릭스 ‘오징어게임’이 최우수 드라마 시리즈 부문을 비롯한 13개 부문 등 총 14개 후보에 올랐으며, 자랑스럽게도 드라마 시리즈 부문 감독상(황동혁) 및 남우주연상(이정재) 등 총 6개 부문에서 수상하며 한국 드라마와 한국 콘텐츠의 새로운 역사를 썼다. 자연스레 한국 최초로 에미상을 받은 황동혁 감독이 이전에 만든 영화가 궁금하여 황 감독의 필모그래피¹⁾를 검색해 보았다. 이름은 낯설었으나 실제로 이미 많은 사람에게 회자하였던, 유명한 영화를 여럿 만들었던, 실로 거장 감독이었다.



네이버에서 황동혁 감독의 필모그래피 검색 결과

메타데이터의 정의

이렇게 미디어 콘텐츠를 경험하다 보면, 콘텐츠에 대한 외적인 정보를 통해 콘텐츠가 주는 즐거움 이외에 새로운 재미를 경험할 수 있도록 콘텐츠에 대한 부가 정보 등이 잘 정리된 것을 확인할 수가 있다. 이러한 정보들은 보통 메타데이터라는 기술적인 용어로 불리며, 대량의 정보 가운데에서 찾고자 하는 정보를 효율적으로 검색하고, 이용할 수 있게 일정한 규칙에 따라 콘텐츠에 대하여 부여되는 데이터를 말한다.

메타데이터의 ‘메타’라는 뜻은 영어의 접두사로, ‘다른 개념으로부터의 추상화’를 가리키며, 후자를 완성하거나 추가하는 데에 쓰인다고 한다.²⁾ 최근에 많이 언급되고 있는 ‘메타버스’에서 사용하는 ‘메타’는 영어 단어 ‘beyond’의 뜻과 유사하게 사용되어, ‘무엇무엇을 뛰어넘어’라는 뜻의 ‘meta’와 ‘세계(universe)’를 나타내는 ‘-verse’와 합해져, ‘가상의 세계’를 나타낸다. 그러나, 메타데이터에서 ‘meta’는 영어 단어인 ‘about’과 유사한 의미로, ‘~에 대해서’라는 뜻으로 쓰인다. 따라서, 메타데이터는 ‘데이터에 대한 데이터’라는 뜻으로 데이터를 설명하는 부가적인 데이터를 지칭한다.

Meta(= beyond) + -verse = Metaverse
Meta(= about) + data = Metadata

의미하는 뜻에 따른 Meta의 다양한 사용

1. 네이버 영화사전, 영화 관계 문헌 혹은 영화 목록을 뜻하는 용어, 일상적으로는 감독, 배우, 제작자 등 영화 관계자들의 고유 영화 목록을 뜻한다
2. 위키백과, ko.wikipedia.org/wiki/%EB%A9%94%ED%83%80

우리가 흔하게 사용하는 메타데이터의 실례(實例)로는 카메라의 메타데이터, 도서의 서지 정보 및 음원 파일의 ID3 태그 정보 등을 꼽을 수 있다. 카메라의 메타데이터는 카메라를 사용하여 사진을 찍을 때마다, 사진을 찍은 날짜, 카메라의 모델 정보, 사용하는 소프트웨어 버전 정보, 사진 해상도 및 크기, 노출값, 조리개값, 초점거리, 화이트밸런스, 대조, 채도, 선명도 등이 자동으로 저장되는 정보를 말한다. 이 정보들은 나중에 사진을 보정하거나 편집할 때 유용하게 사용할 수 있는 값들을 위주로 저장된다고 할 수 있다.

음원 파일의 태그 정보도 마찬가지이다. 음원 파일의 ID3 태그는 미디어 플레이어에서 음원 파일을 재생할 때 해당 음원에 대한 각종 정보를, 이를테면 가수, 작곡자, 작사가, 장르, 트랙번호, 앨범명 및 가사 정보 등을 제공한다.

도서의 서지 정보와 같은 메타데이터는 흔히 말하는 ISBN(International Standard Book Number, 국제표준도서번호)이나 ISSN(International Standard Serial Number, 국제표준간행물번호) 등이 될 수도 있고, 작가, 출판사, 번역자 그리고 최근에는 평점 등도 메타데이터 내에 정보로 담고 있는 예도 있다.

우리가 도서관이나 온라인 서점에서 책을 찾을 때 ‘제목’과 함께 가장 많이 사용하는 ‘질의어’, ‘검색용 키워드’가 메타데이터 정보들인 것이다. 종합적으로 보면, 메타데이터는 해당 메타데이터가 설명하고 있는 데이터나 미디어 콘텐츠 파일을 잘 설명하거나, 잘 찾을 수 있게 도와주며, 또는 잘 관리할 수 있게 하는 등 미디어 콘텐츠의 활용성을 높이는 데에 직간접적으로 도움을 준다고 할 수 있다.

메타데이터는 재사용성과 이용의 효율성을 고려하여 계층구조로 구성되어 있으며, 많은 기관이나 단체에서 상호운용성을 확보하기 위해 해당 분야에 알맞은 표준을 정의하고 그것에 따라 메타데이터를 개발하여 사용해왔다. 따라서, 메타데이터는 정의하는 단체의 입장과 단체의 설립 목적 등에 따라 여러 가지 방법 및 다른 단어들을 사용하여 정의하고 있지만, 대부분 비슷한 의미를 지닌다. 1996년 더블린코어 메타데이터 워크숍에서 논의한 메타데이터는 ‘데이터에 관한 데이터’로서 정보자원의 명세, 탐색, 관리 및 운용 전반을 지원하기 위한 데이터라고 정의하였다. 앞서 예로 들었던 것들 가운데, ‘명세’는 음원 파일을 재생하는 미디어 플레이어에 제공하는 정보가 사례가 될 것이고, ‘탐색’은 도서 검색을 할 때 키워드로 사용하는 메타데이터 등을 들 수가 있다. 또한 ‘관리’는 카메라에서 획득한 영상 정보에 포함되는 메타데이터를 활용하여 관리를 쉽게 할 것이다.

즉, 메타데이터의 존재 자체가 실제 미디어 데이터를 좀 더 효과적으로 사용할 수 있도록 지원하는 데이터라는 것이며, 메타데이터는 없어도 상관없으나 있으면 본래 데이터의 사용성이 한층 증가하게 된다는 것을 확인할 수가 있다. 미국 정보 기술 연구 및 자문회사인 가트너에서도 메타데이터는 특정 정보 자산의 여러 측면을 기술하여, 정보가 생성되고, 유통되고 또한 관리하는 시점까지, 정보의 수명 주기 전반에 걸쳐 그 유용성을 개선하기 위한 정보라고 정의하고 있다.³⁾

Gartner Glossary > Information Technology Glossary > M > Metadata

Metadata

Metadata is information that describes various facets of an information asset to improve its usability throughout its life cycle. It is metadata that turns information into an asset. Generally speaking, the more valuable the information asset, the more critical it is to manage the metadata about it, because it is the metadata definition that provides the understanding that unlocks the value of data.

가트너가 정의하는 메타데이터

3. 가트너, www.gartner.com/en/information-technology/glossary/metadata

메타데이터의 분류

앞서, 대표적인 메타데이터 사례를 살펴보면, 메타데이터가 포함하고 있는 정보를 잠깐 확인하였는데, 메타데이터가 포함하고 있는 정보에 따라 메타데이터는 다음과 같이 분류할 수 있다. 먼저, 데이터 식별에 필요한 식별자, 제목, 저자 등의 정보와 같은 유일한 값을 다루는 '식별정보'를 꼽을 수 있다. 식별정보는 말 그대로 한 데이터와 다른 데이터를 구분할 수 있는 유일한 정보를 일컬으므로, 데이터를 검색할 때 주로 질의어로 활용되는 메타데이터가 될 것이다. 그리고, 데이터의 파일 사이즈, 해상도, 파일 포맷, 그리고 특히 카메라로 획득한 영상 정보의 메타데이터 중 노출값, 조리개값, ISO값 등 물리적인 요소에 해당하는 메타데이터는 '특성정보'라고 할 수 있다. 특성정보는 일반적으로 자동으로 생성되는 값들이 대부분이다. 사진을 찍어 영상을 생성하거나, 음원 파일의 마스터링을 마치게 되면 자동으로 생성되는 정보 등이 메타데이터 중 특성정보라고 할 수 있다. 데이터가 포함하고 있는 정보를 빠르게 훑어볼 수 있도록 전체 내용에 대한 요약 정보를 메타데이터에 포함하기도 하는데(예를 들어, 논문의 초록 정보 등), 이러한 정보는 '내용서술정보'라고 말한다. 본래 데이터를 완벽하게 이해할 수 있어야 데이터의 내용을 전체적으로 간추릴 수 있으므로 내용서술정보도 일반적으로 수동으로 입력을 해야 한다.

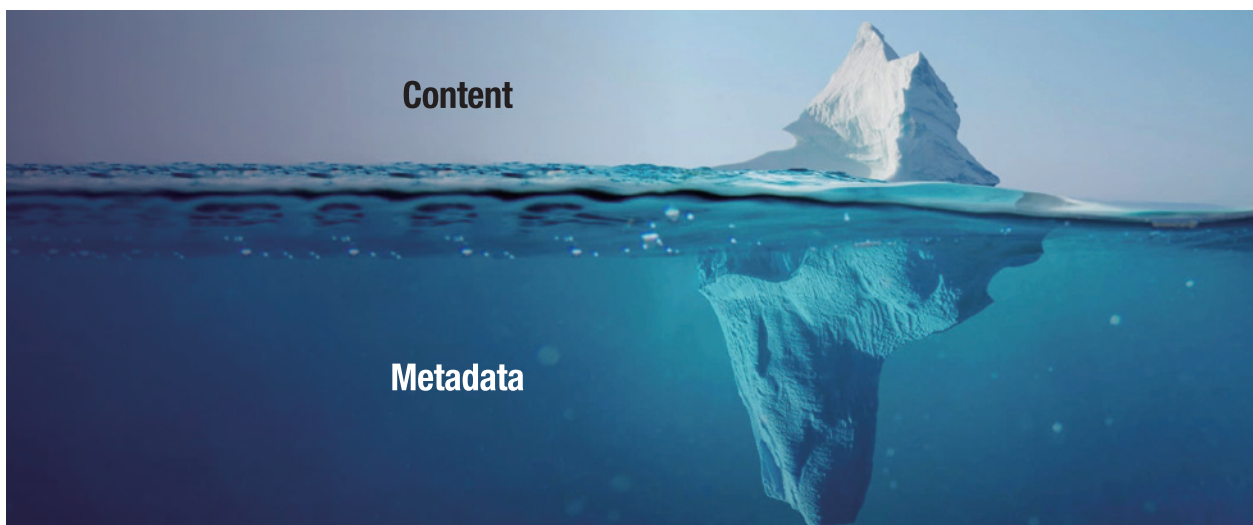
또한, 많은 데이터 가운데에서 원하는 데이터나 정보에 빠르게 접근할 수 있도록 데이터 간에 구분을 할 수 있는 제목, 주제, 키워드나 저자 등의 정보를 '검색정보'라고 하는데, 식별정보와 유사점을 많이 가지고 있다. 식별정보나 검색정보도 미디어가 가지고 있는 유일한 정보를 꼽아 입력해야 하므로 수동으로 추가하는 경우가 대부분이다. 마지막으로 데이터 관리에 필요한 정보, 예를 들어 생성날짜, 수정날짜나 버전 정보 등의 정보를 메타데이터의 '관리정보'라고 부른다. 이렇게 정보의 내용에 따라 분류한 메타데이터의 내용을 보면, 데이터나 미디어 콘텐츠가 생성될 때 자동으로 만들어지는 정보가 있는가 하면, 수동으로 만들어줘야 하는 정보가 있다. 일반적으로 수동으로 만들어야 하는 정보는 사람의 노동력이 수반되는 것이다.

그래서 메타데이터를 작성하는 방법에 따라 다음과 같이 분류할 수도 있다. 첫 번째 자동으로 추출이 가능한 메타데이터로서 미디어 콘텐츠의 경우 카메라의 동작, 밝기 정도 등 가공되지 않은 특징을 표현하는 정보를 '내용 의존적인(Content dependent)' 메타데이터라고 한다. 다음으로는 자동으로 추출하기 어려우나 일반적으로 직관적인 정보를 수동으로 입력해야 하는 메타데이터를 꼽을 수가 있는데, 예를 들면, 콘텐츠 제목, 날짜, 제작자, 제작시간 등 콘텐츠의 전체적인 특징을 표현하는 '내용 독립적인(Content independent)' 메타데이터가 바로 그것이다. 마지막으로 메타데이터를 수동으로 입력하는 작성자의 직관과 경험 정도에 따라 내용에 차이가 발생할 수 있는 '내용 설명적인(Content descriptive)' 메타데이터가 있을 수 있다. 이 정보는 미디어 콘텐츠의 순서, 내용 등 관련된 특징을 표현하는 정보라서 작성자의 주관적인 면이 반영될 수가 있다.

이처럼, 메타데이터는 자동으로 또는 수동으로 작성할 수가 있는데, 수동으로 작성을 하더라도 입력하는 작성자의 경험이나 노하우, 또는 해당 콘텐츠와 관련한 배경지식 등으로 인해 동일한 미디어 콘텐츠에 대하여 서로 다른 품질의 메타데이터가 입력될 수 있다는 가능성이 존재한다. 이렇게 보면, 메타데이터 정보를 구축하는 것만이 능사가 아니다. 전문지식을 갖춘 유경험자 또는 메타데이터를 작성하는 일관된 지침 등이 존재해야만, 추후 활용 가능성이 크며, 내용에 일관성 및 통일성을 갖고 있는 메타데이터로 사용할 수 있을 것이다.

미디어 콘텐츠와 메타데이터

ICT(Information Communication Technology)의 발달과 융합을 통하여 2010년대 초반, 아날로그 방송이 종료되고, 디지털 방송이 전국화되었으며, 유튜브, 넷플릭스 등 OTT(Over-the-Top) 서비스의 급부상으로 인해, 국내는 물론 미디어 콘텐츠의 제작(생산), 송출(유통) 및 소비 환경은 큰 변화를 맞이하고 있다. 기존 신문사, 방송사들을 대표하는 '레거시 미디어'의 위기라는 것이 큰 변화를 대표하는 흔한 표현이 된 것이다. 그럼에도 불구하고, 방송사를 비롯한 레거시 미디어들은 급박하게 패러다임이 전환되는 국내외 방송 환경 속에서 내부적으로 자강 정책을 통해 변화하는 시류에 적극적으로 대응 중이다. 방송 콘텐츠 기획, 제작, 송출 및 관리의 워크플로우를 대용량 저장장치나 네트워크 기반의 온라인 시스템으로 전환하여, 방송사 또는 제작사 내 별도로 구축되어 있는 각 시스템에 대한 유기적이고 통합적인 관리에 대한 필요성이 함께 증대되고 있는 것이 사실이다.



특히, 미디어 콘텐츠는 그 시대를 대변하는 기록이라는 역사적, 교육적 가치를 보유하고 있어서 메타데이터를 추가하여 잘 보존하고 새로운 활용처에 적용해야 할 것이다. 미디어 콘텐츠는 인쇄 매체와는 구별되는 구체성과 입체성을 보유하고 있으며, 시대사나 특정 부분에 대한 부문사 연구에 필수적인 학술 자료로서 역사적 가치를 가지고 있다. 또한 우리 사회의 한 단면을 보여주는 효과를 가지고 있어서, 다른 시대와 비교 관찰을 할 수 있는 문화유산으로서의 교육적 가치 또한, 가지고 있다.⁴⁾

미디어 콘텐츠는 저장 매체가 단순히 파일로만 국한되지 않고, 전통적인 필름이나 자기테이프 등의 다양한 형태로 존재해 왔고, 같은 내용을 담고 있는 미디어 콘텐츠라고 하더라도 쓰임새에 따라 방송본, 촬영본, 편집본과 같은 다양한 구분이 존재한다. 또한, 근 몇 년간에는 OTT 및 SNS(Social Network Service) 등을 통해 미디어 콘텐츠의 제작자와 사용자의 구분 없이 수많은 미디어 콘텐츠가 실시간으로 만들어지고 소비되는 환경으로 변화하는 중인 것이다. 따라서, 방송사, 콘텐츠 사업자 및 1인 미디어 제작자들은 미디어 콘텐츠의 재사용성과 효율적인 활용성을 위하여 미디어 콘텐츠를 체계적으로 관리하고 안전하게 보존하기 위해 미디어 아카이브 시스템을 구축해서 활용하고 있다. 하지만 아카이브 시스템 내에 미디어 콘텐츠가 시시각각 쌓여갈수록, 필요한 미디어 콘텐츠를 적시, 적기에 찾아내고 활용하기 위해서는 반드시 미디어 콘텐츠에 대한 메타데이터를 적절하게 제작하고 이에 대한 사용 방법이나 지침 등을 마련해 놓아야만 한다.

따라서 메타데이터는 본래 미디어 콘텐츠가 가지고 있는 구조화된 정보를 분석하고 분류하여, 부가 정보를 추가하기 위해 데이터에 따라가는 정보로서 본래 정보를 자산화한다는 데에 의의가 있다. 즉, 정보 자체에 중요성을 더 부여하기 위해서 굳이 부가적인 정보를 일일이 만들어 정보에 엮어서 보관하는 것이다. 그러다 보니, 최초에 이러한 정보는 필요할 것이라고 예상하고 추가해 놓은 메타데이터가 실제로는 쓰이지 않아서 유명무실해지는 정보도 있을 것이고, 시대와 흐름에 따라 새롭게 메타데이터로서 정의를 해야만 할 정보들도 있을 것으로 생각한다.

앞으로, 총 3회에 걸쳐 국내외 표준화 사례를 통해 메타데이터를 어떻게 활용해 왔으며, 현재까지도 메타데이터의 활용성을 높이기 위해 국가 연구개발 과제로서 어떤 연구가 진행되고 있는지, 그리고 마지막으로 이렇게 메타데이터를 잘 구축해 놓으면, 활용 가능성이 얼마나 커지며 어떻게 미디어 콘텐츠를 부가적으로 경험할 수 있는지를 설명할 예정이다. 

4. 김성철, '지상파 방송사 영상 콘텐츠 보존과 디지털 영상 아카이브 연구' 서강대 언론대학원 석사학위논문, 2006