

## 미디어 창작을 위한 비디오 아카이브 키워드기반 내용 검색 서비스 요구사항 분석

\*정병희 \*\*박완 \*\*\*이윤성 \*\*\*\*이하주 \*\*\*\*\*김산성

한국방송공사(KBS)

\*bhjung@kbs.co.kr

Analysis of Keyword-based Content Search Service Requirements  
in Video Archive for Media Creation

\*Byunghee Jung \*\*Wan, Park \*\*\*Yunseong, Lee \*\*\*\*Hajoo, Lee \*\*\*\*\*Sansung Kim

Korean Broadcasting System

## 요약

방대한 분량의 콘텐츠 홍수 속에서 원하는 소재를 찾기 위해 콘텐츠 내용을 검색할 수 있는 효과적인 방법이 지원되는 것은 창작을 자유롭게 하고, 콘텐츠 활용도를 높이기 위해 매우 중요하다. KBS 바다 서비스의 경우 분류체계 방법을 사용하고 있으나, 최근 딥러닝을 이용한 인공지능 기술의 발전으로 콘텐츠의 내용을 인공지능 기술로 태깅하고, 태깅된 텍스트 정보를 이용하여 검색할 수 있는 기술 개발이 활발히 수행되고, 국가적으로도 해당 기술을 지원하고 있다.

본 논문에서는 이러한 기술 개발의 선행 요소인 방송사의 제작과정에서 요구되는 동영상 소재 콘텐츠 검색의 요구사항을 KBS 비디오 아카이브 검색 키워드 실제 사용 데이터를 이용하여 분석하였다. 약 1,000여건의 검색 키워드 분석과 이용자와 운영자의 응답 내용을 고찰한 결과, 특정 키워드에 집중하여 검색할 수 있도록 보완하여 주는 것이 필요함을 알아내었다. 또한, 검색 범위를 효과적으로 축소하여 검색을 손쉽게 빠르게 할 수 있는 방법을 고찰하였다. 본 논문에서는 미디어 창작에서 필요한 소재 콘텐츠를 찾기 위해 연구 개발해야 할 미디어 속성 추출 기술의 방향성을 제시하였다.

## 1. 서론

한국방송공사는 방송 제작을 위한 다양한 콘텐츠를 제공하며, 방송 유산으로서의 콘텐츠 가치를 보존하고 극대화하는 목표를 가지고 'KBS 비디오 아카이브 시스템'(이후 'KBS 비디오 아카이브' 혹은 '비디오 아카이브'로 표시)을 구축하여 운영하고 있다. 2000년대에는 테이프 등의 형태로 방송 콘텐츠를 보관하고 이를 디지털로 검색하는 시스템을 개발하여 운영하였으며, 2011년 본격적인 디지털 시대를 맞아 콘텐츠를 모두 디지털화하고, 검색 기능 등을 일원화하여 관리하는 시스템을 도입하고 현재까지 발전시켜 오고 있다. 1단계 서비스를 개시한 2011년 이후 현재까지 본사의 방송본 프로그램을 디지털화하여 저장하고 활용하였으며, 서비스 대상을 해외 계열사와 지역총국의 제작에도 지속 확대하고 있다. KBS 비디오 아카이브는 콘텐츠 가치를 극대화하기 위해 인제스트된 콘텐츠의 품질을 검증하는 절차를 거치며, 활용을 극대화하기 위해 KBS에서 제정한 메타데이터 표준화에 근거한 메타데이터 입력 작업을 수행하고 있다. 이 모든 작업은 KBS 내 전담 부서에서 제정한 작업 규칙

에 따라 수행한다 [1].

2022년 3월, 한국방송공사는 전 국민 대상으로 창작자에 대한 콘텐츠의 자유로운 활용이 가능한 무료 콘텐츠 클립 공개 서비스를 오픈하였다. 'KBS 바다'라고 명명한 서비스는 콘텐츠 제작, 학습, 교육, 연구에 짧은 고품질 소재 콘텐츠를 활용할 수 있도록 하는 서비스이다. 이 서비스에도 이용자들이 콘텐츠를 쉽게 접근할 수 있도록 다면 분류체계를 이용하여 키워드 중심으로 공개된 콘텐츠를 접근할 수 있도록 하고 있다 [2].

방대한 분량의 콘텐츠 홍수 속에서 원하는 소재를 찾기 위해 콘텐츠 내용을 효과적으로 검색하는 방법이 지원되는 것은 창작을 자유롭게 하고, 콘텐츠 활용도를 높이기 위해 매우 중요하다. 전술한 KBS 바다 서비스의 경우 분류체계 방법을 사용하고 있으나, 최근 딥러닝을 이용한 인공지능 기술의 발전으로 콘텐츠의 내용을 인공지능 기술로 태깅하고, 태깅된 텍스트 정보를 이용하여 검색할 수 있는 기술 개발이 활발히 수행되고, 국가적으로도 해당 기술에 대한 지원 중요성이 제기되고 있다 [3].

본 논문에서는 이러한 기술 개발의 선행 요소인 방송사의 제작과정에서 요구되는 동영상 소재 콘텐츠 검색의 요구사항을 KBS 비디오 아카이브 활용 사례를 분석하여 고찰하고자 한다. 본 논문의 구성은 다음과 같다. 2절에서는 비디오 아카이브 사용자들이 소재를 찾기 위해 일정 기간 입력한 검색 키워드를 분석하여 어떤 종류의 내용기반 메타데이터를 태깅하여 제공하는 것이 원하는 영상을 찾는 데 도움이 되는지 분석한다. 3절에서는 KBS 비디오 아카이브 사용자 및 운영자를 대상으로 비디오 아카이브 활용 시 요구사항들을 정리하여 기술한다. 4절에서는 이를 기반으로 전체적인 요구사항 분석 결과를 고찰한다.

본 논문에서 분석한 요구사항은 추후 미디어 창작에서 필요한 소재 콘텐츠를 찾기 위해 연구 개발해야 할 미디어 속성 추출 기술의 방향성으로 제시하고자 한다.

## 2. 비디오 아카이브의 이용자 검색 키워드 분석 실험

본 절에서는 KBS 비디오 아카이브에서 1주간 (2021.5.31. ~ 2021.6.6.) 사용된 검색 키워드 총 5,000여건 중 1,000여건을 대상으로 분석하였다. 대상 검색 키워드의 사전적 의미와 형식에 의해 분류해 보면, 표 1.의 종류로 구분되었다. 검색 키워드에는 고유 명사가 많이 사용되었는데, 여기에는 인물명(주요인사, 출연자명, 가수명, 그룹명 등)과 특정 지역명(국가, 도시, 자연지명 등), 랜드마크명, 이벤트명 등이 포함되었다. 고유 명사는 보통 명사처럼 여러 키워드들을 조합하거나 상황에 따라 의미가 다르게 인식되지 않고, 검색 대상을 명확히 지칭할 수 있어 키워드 기반의 검색에 매우 효과적이다.

보통 명사의 경우에는 검색하고자 하는 콘텐츠의 관심 주제 내용을 입력하는 경우가 다수였으며, 사물이나 카메라 기법 등 매우 다양한 내용으로 입력되었다. 또한, 보통 명사의 경우에는 고유 명사를 대상으로 하여 콘텐츠의 내용이나 형식을 더욱 특정하기 위해 고유 명사와 같이 검색 키워드로 병기하여 검색 의도를 명확히 하는 경우가 약 30%로 나타났다.

표 1. 검색 키워드 살사용 사례 분석

구분	비율	실례
프로그램명	24.5% (251)	뮤직뱅크, 불후의 명공, 아침마당
이벤트명	3.3% (34)	평창올림픽, 2002월드컵
지역명	7.4% (76)	중국, 싱가포르, 하와이, 북한산
인물명	15.8% (162)	박세리, 방탄소년단, 정주영
주제 (보통명사)	39.3% (403)	검은콩, 시위, 타임랩스, 해수욕장 인파, 옛날 골목, 컬링, 단풍
식별ID	6.3% (65)	512692150 (KBS 비디오아카이브ID)
장르명	0.5% (5)	뉴스, 다큐, 시사기획
랜드마크명	2.9% (30)	부산역, 경포대, 대청댐
합계	1,026	

## 3. 비디오 아카이브의 이용자 및 운영자 응답 요구사항

KBS 비디오 아카이브의 이용자 및 운영자를 대상으로 문익한 비디오 아카이브 콘텐츠 검색의 요구사항은 다음과 같다.

사용자들은 비디오 아카이브의 검색 키워드 입력 방식에 의한 콘텐츠 검색 방식에 적응한 형태를 보인다. 사용자는 원하는 콘텐츠 검색 시 검색 키워드 입력으로 빠르게 1차 결과를 받아보는 것에 불편을 느끼지 못하며, 이미지 쿼리 입력 방식 등 새로운 검색 방식에 대한 요구는 운영기간 없었던 것으로 응답되었다. (해당 응답은 비디오 아카이브를 담당해 온 관리 전담 부서의 담당자 5명으로 구성된 심층 면담의 결과임)

단, 키워드 입력에 의한 검색 시, 상세 조건 입력, 검색 결과 내 날짜 등의 추가 조건 입력을 통한 재검색 등 1차 결과 내에서 검색 결과 범위를 축소하는 검색 편의 기능을 추가하는 요구사항이 있는 것으로 응답되었다. 또한, 비디오 아카이브는 주로 영상 콘텐츠를 대상으로 하므로 타임라인 기반으로 검색 키워드와 가장 일치하는 부분을 표시하거나, 일치도를 표시하여 영상 내에서 검색 결과 위치를 빠르게 찾아갈 수 있게 하는 기능에 대한 추가 요구사항도 파악되었다. 이와 같은 응답에 의한 이용자 및 운영자가 요구하는 키워드 입력 기반의 검색 방식을 도식화하면 그림 1.과 같다.

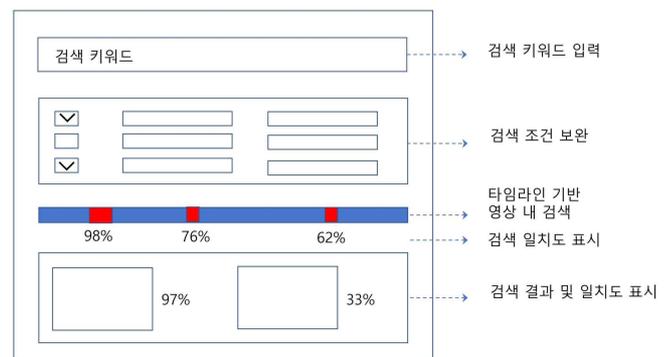


그림 1. 키워드 입력 기반의 검색 방식 요구사항

검색 키워드를 보완하는 방식으로 검색 편의 기능을 추가하는 방법에 대한 요구사항도 파악되었다. 특히, 인공지능 기술을 이용해 콘텐츠 내용에 대한 추가 속성 정보를 검색 가능하도록 제공할 시, 콘텐츠 검색 정확도를 높일 수 있다는 응답 결과가 있었다. 자동 추출 가능한 검색어 키워드 종류와 해당 키워드의 예상 활용도에 따른 세부 요구사항은 다음과 같다.

- 인물 : 인물명은 가장 기본적인 검색 접근임
  - 수동 입력으로 확보되지 못한 인물 검색어 보완
  - 출연 당시는 유명하지 않았던 인물의 재조명 용도로 사용 가능
- 사물 : 제작 및 판매용으로 활용도 높은 사물 대상이 유용함
  - 일반 사물 : 지폐, 약, 식기, 핸드폰, 컴퓨터, 장독대 등
  - 음식 : 내부 제작 이용 및 외부 판매 빈도 높은 대상 위주
  - 특정 상품(제품/상표) : 광고 및 협찬 관련 신사업 모색 가능
  - 동물, 식물 : 내부 제작 이용 및 외부 판매 빈도 높은 대상
- 맥락 : 전형적 동작과 상황이 대상이 되어야 함
  - 인서트용으로 사용되는 전형적 동작과 상황 :

음료를 먹는 / 기지개를 펴는 / 시위하는 / 예방 접종하는 / 줄 서 있는 / 부채를 부치는 사람들 등

- 촬영기법 : 검색 요구가 많은 특수 촬영 기법 위주
  - 드론, 항공, 부감, 양각, 파노라마, 타임랩스, 미속, 고속 등
- 배경·풍경 및 기상현상 : 제작 및 판매용으로 활용도 높은 현상
  - 일반적 배경 : 설산, 바다, 차도, 길거리, 산지, 남극, 북극, 인파, 실내외, 병원 등
  - 기상현상 : 일출, 일몰, 태풍, 토네이도, 오로라, 백야, 극야, 구름, 폭우, 번개, 개기일식
  - 전형적 위치를 연상하는 그림 : 경찰청, 감옥, 학교, 달동네, 병원, 사무실)
- 국가, 위치, 랜드마크 : 특정 국가에서 일어난 특정 상황 또는 전형적 상황에 대한 검색 빈도 높음
  - 미안마 시위, 베니스 페스티벌, 미 의회 전경, 에펠탑의 관광객, 머드 축제
- 텍스트 관련 메타데이터 : STT / OCR / 자동번역
  - 자동 음성 인식 (STT) / 화면 내 글자 및 자막 인식 (OCR)
  - STT와 OCR을 통해 얻어진 텍스트 데이터의 다국어 번역

#### 4. 방송 제작 활용을 위한 비디오 아카이브 요구사항 고찰

본 논문에서는 KBS 비디오 아카이브 이용자의 실사용 검색 키워드에 대한 분석과 응답을 통한 심층 분석을 수행하였다. 분석 결과 이용자 및 운영자들은 검색 방법에 있어 검색 키워드 입력에 의한 방법을 가장 기본적인 방법으로 인식하였고, 적응하는 형태를 보였다. 특히 키워드 기반의 검색 편의 기능을 추가하는 요구사항이 많았고, 키워드 입력 방식 이외의 입력(예를 들면, 이미지나 스케치 입력 등) 방식 개선에 대한 요구는 없었다.

실사용 검색 키워드는 프로그램명 24.5%, 고유 명사가 29.5%를 차지해 검색 대상을 단독으로 특정할 수 있는 키워드 위주로 약 55% 정도의 검색이 이루어졌다. 고유 명사의 경우는 인물명, 지역명, 이벤트명, 랜드마크명의 순으로 많이 사용되었는데, 이는 인공지능 기술을 이용해 속성을 추출할 경우 많이 사용되는 고유 명사 위주로 추출하면 방대한 분량의 콘텐츠 검색에 도움을 줄 수 있다는 것을 의미한다. 또한, 고유 명사 위주의 자동 속성 추출 기능을 연구개발할 시, 이용자 및 운영자가 응답한 활용이 예상되는 키워드 위주로 우선 진행하는 것이 활용도를 높일 수 있는 접근 방법이다.

검색의 40%에 해당하는 보통 명사를 검색 키워드로 사용하는 경우, 1) 내용의 주제와 관련된 사물, 2) 장소이거나, 3) 주제어(추상명사)인 경우가 많았다. 특히 사물 혹은 장소 (자연, 특정 지형을 일컫는 보통명사 등의 경우, 인공지능 기술을 이용한 속성 추출의 대상이 될 수 있어 자동화 가능한 속성이 될 수 있다.

고유 명사를 검색 키워드로 사용하는 경우 주제어를 병기하여 검색을 더욱 빠르게 시도하는 경우가 50%이상을 차지하였다. 특히, 인물 혹은 음악 퍼포먼스 그룹의 경우, “방탄소년단 무대”, “BTS 연설” 등과 같이 해당 인물 혹은 그룹이 출연하는 상황에 대한 키워드를 병기하여 검색을 빠르게 하기 위한 목적으로 여러 개의 키워드를 나열하였다. 랜드마크명은 주로 해당 랜드마크가 독자적으로 활용되는 경우가 많아 단독으로 사용되는 경우 많았지만, 지역명의 경우는 나라, 도시, 산과 강을

중심으로한 지역 등 범위가 다양하여 지역명과 기타 주제 키워드가 같이 병기되어 검색되는 경우가 많았다.

방송 제작의 특이사항으로 숙련된 이용자의 경우 검색 키워드 입력 대신, 익숙한 코드나 키 값을 사용하는 경우도 6%로 상당수 나타났다.

#### 5. 결론

방대한 분량의 콘텐츠 홍수 속에서 원하는 소재를 찾기 위해서는 콘텐츠 내용을 검색할 수 효과적인 방법이 지원되는 것은 창작을 자유롭게 하고, 콘텐츠 활용도를 높이기 위해 매우 중요하다. KBS 바다 서비스의 경우 분류체계 방법을 사용하고 있으나, 최근 딥러닝을 이용한 인공지능 기술의 발전으로 콘텐츠의 내용을 인공지능 기술로 태깅하고, 태깅된 텍스트 정보를 이용하여 검색할 수 있는 기술 개발이 활발히 수행되고, 국가적으로도 해당 기술을 지원하고 있다.

본 논문에서는 이러한 기술 개발의 선행 요소인 방송사의 제작과정에서 요구되는 동영상 소재 콘텐츠 검색의 요구사항을 KBS 비디오 아카이브 검색 키워드 실사용 사례 데이터를 이용하여 분석하였다. 약 1,000여건의 검색 키워드 분석과 이용자와 운영자의 응답 내용을 고찰한 결과, 특정 키워드에 집중하여 검색할 수 있도록 보완하여 주는 것이 필요함을 알아내었다. 활용도가 높은 키워드로는 인물명, 지역명, 이벤트명, 랜드마크명의 순이었다. 또한, 2개 이상의 키워드를 병기하여 입력하는 것이 검색을 쉽고 빠르게 할 수 있음이 입증되어, 향후, 자동 속성 추출을 통한 검색 입력 시 병기하는 경우에 대한 고려가 필요하다.

본 논문의 경우 소재 콘텐츠(동영상 등)를 검색하는 경우를 대상으로 분석을 실시하였다. 내용이 특정되는 10초 미만의 클립이나, 이미지의 경우에는 넓은 범위의 주제어를 입력하지 않아도 되므로, 짧은 길이 소재 콘텐츠나 이미지 검색의 경우에 대해서는 추가 고찰이 필요하다. 2절의 사례에서 주제의 경우는 다양한 범위의 키워드 입력 사례를 보인다. 따라서 향후에는 검색 키워드와 주제어의 연관성 혹은 주제어의 범위 한정에 관한 추가 연구가 필요하다.

#### 참고 문헌

- [1] KBS 아카이브 소개, <http://kbsarchive.com>
- [2] KBS 바다, <http://bada.kbs.co.kr>
- [3] 조용성 외, “미디어와 AI기술: 미디어 지능화”, 전자통신동향분석 35권 제5호, 2020년 10월

#### Acknowledgement

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00852, 지능적 미디어 속성 추출 및 공유 기술 개발).