

# 동영상 스트리밍 서비스에 기반한 실시간 인터랙티브 자막 재구성 서비스\*

노현<sup>o</sup> 박상근

경희대학교 소프트웨어융합대학 소프트웨어융합학과  
yeshyun@khu.ac.kr, sk.park@khu.ac.kr

## Real-time Interactive Automatic Caption Reconstruction On Video Streaming Services

Hyun Roh<sup>o</sup> Sangkeun Park

Department of Software Convergence, KyungHee University

### 요 약

본 연구는 COVID-19 로 인한 실내 문화생활 콘텐츠 수요 증가와 함께 동영상 미디어 콘텐츠의 중요성이 커지고 있는 상황에서, 유튜브와 같은 동영상 플랫폼의 자동 자막 기능을 개선하여 영상 시청 경험을 향상시키는 서비스를 개발하였다. 이 서비스는 영상 아래에 자막을 배치하여 시청 중에 영상을 가리지 않도록 하고, 자연어 처리 기술을 사용하여 자막을 문장 형태로 재구성하며, 이전과 다음 자막을 함께 제공하여 문맥 파악을 용이하게 한다. 또한 사용자가 자막의 스타일을 커스터마이징할 수 있으며, 감정 분석 결과에 따라 자막 스타일을 동적으로 변경하여 가독성을 높이고 생동감 있는 시청 경험을 제공한다.

### 1. 서 론

최근 COVID-19 이슈로 인해 사람들의 실내 문화생활 콘텐츠 소비가 증가함에 따라 다양한 영상 미디어 콘텐츠의 수요가 함께 증가하고 있다. 전체 동영상 OTT 서비스 시장으로 보았을 때 2023년에는 약 1653 억 달러의 시장 규모로, 2021년 대비 약 25% 상승했다. 넷플릭스, 유튜브 등 글로벌 기업을 중심으로 빠른 속도로 성장하고 있는 만큼 이러한 OTT 시장의 성장세는 3년 동안 약 1.3 배 이상 성장할 것으로 예상된다. [1]

전 세계로 동영상 서비스 플랫폼이 확산됨에 따라서, 각 나라에 맞는 언어로 번역이 되어야 하고 시청 경험을 높이기 위해 자막 사용이 보편화되었다. 동영상 내의 자막은 언어적 측면뿐만 아니라 그래픽 요소의 활용 측면에서도 중요하게 다뤄지고 있다. 서체, 색상, 움직임, 화면 배치 등을 통해 자막의 효과를 극대화하여 사용자에게 생동감 있는 경험을 제공한다. [2]

하지만 영상을 업로드하는 편집자가 자막을 제공하지 않으면 시청자는 서비스에서 자체적으로 제공하는 자동 자막을 사용하게 되는데 이때 대부분의 플랫폼에서 영상을 가리는 문제가 발생한다. 또한 현재 자막만 제공하고,

실시간으로 자막을 생성하는 플랫폼에서는 문장별로 제공되지 않아 문맥을 파악하기 어려운 경우가 있다. 더불어 정적인 스타일을 제공하여 시청자는 자동 자막을 사용할 때 생동감 있는 경험을 할 수 없다.

이에 본 연구에서는 동영상 플랫폼 중 대한민국 인구의 약 81%가 사용하고 있는 유튜브(Youtube) [3]의 자동 자막을 활용한 보기 편리한 유튜브 자동 자막 제공 서비스를 개발했다. 시청자가 유튜브 자동 자막 기능을 사용할 때 자막을 영상 밑에 배치하고, 자연어 처리 기술을 활용하여 자막을 문장으로 재구성한다. 이후 이전 자막과 다음 자막을 함께 제공하여 문맥 파악에 용이하게 하고, 스타일 변경 기능을 제공하여 생동감 있는 자막 경험을 제공하도록 구현했으며 사용자 스타디를 통해 본 연구의 아이디어 효과를 검증했다.

### 2. 관련연구

#### 2.1 자막 배치 수정 및 문장 제공

적절한 자막 배치를 통해 시청자의 몰입도와 문맥 이해도를 높이기 위한 다양한 연구가 수행되었다. 영상 내의 자막을 발화자의 얼굴 주변과 같은 정보성이 낮은 공간에 자동으로 배치하는 시스템을 개발했으나, 자막이 영상을 가리는 문제는 여전히 해결되지 않았다. [4] Language Reactor 서비스에서는 유튜브 자동 자막을 자동으로 번역해 주는 기능을 제공한다.

\* 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2023년도 SW 중심대학사업의 결과로 수행되었음 (2023-0-00042)

자막을 보기 쉽도록 현재 자막과 다음 자막을 함께 제공하지만, 문장으로 재구성하는 것이 아닌 기존 유튜브에서 제공하는 자동 자막을 이어 붙인 한계점이 존재한다. 또한 자막의 위치가 영상 내에 위치하여 영상을 가리는 한계점이 존재한다.[5]

본 연구에서는 자연어 처리 기술을 활용하여 자막을 문장으로 재구성한 뒤, 이전과 다음 자막을 함께 제공한다. 이후 영상 시청을 방해하지 않는 공간인 영상 하단에 자막을 배치하여, 기존 연구들의 한계점을 해결하고 시청자의 몰입도와 문맥 파악 경험을 높인다.

### 2.2 자막 커스터마이징

자막의 스타일을 변경하여 가독성을 높이기 위한 연구가 수행되었다. 영상에서 한국어 음성을 추출한 후 ACNN 모델을 활용하여 감정을 분석했다. 감정에 따라 그에 상응하는 자막에 색상을 입혀 자동으로 자막을 생성해 주는 모델을 제안했으나, 편집자만이 수정이 가능하고 사용자가 변경할 수 있는 기능을 제공하지는 않은 한계점이 존재했다 [6]. Mehrabian 의 PAD 감정 분류 모형을 활용하여 음악 가사의 감정을 분석했다. 분석 결과에 따라서, 긍정의 감정은 Sans-Serif, 부정의 감정은 Hand Writing 폰트로 표현하며 자막을 자동으로 생성해 주는 서비스를 연구했지만, 개인의 선호와 주관을 고려하지 못한 한계점이 있다.[7]

본 연구에서는 시청자가 자막의 색상, 폰트, 배경을 변경할 수 있는 커스터마이징 기능을 고려한다. 사용자의 주관과 선호에 따라서 자막의 가독성을 높이고 즐거운 영상 시청 경험을 제공하고자 하였다. 또한 감정 분석을 통해 도출된 긍정과 부정의 결과에 따라서 자막의 스타일을 변경하여, 보다 생동감 있는 영상 시청 경험을 제공할 수 있다.

### 3. 설계 및 구현

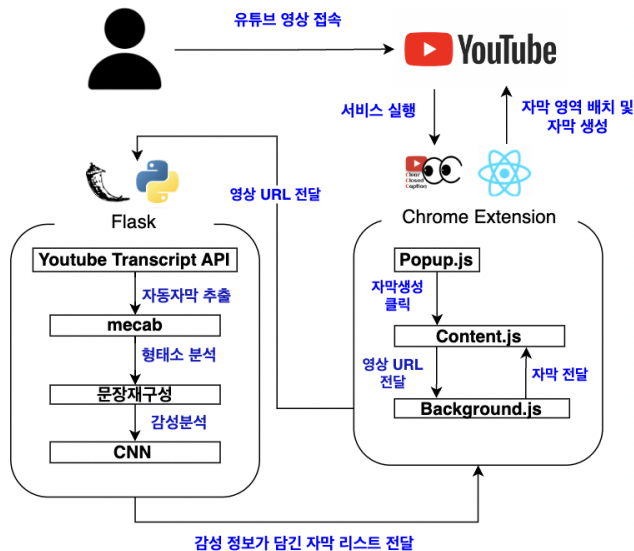
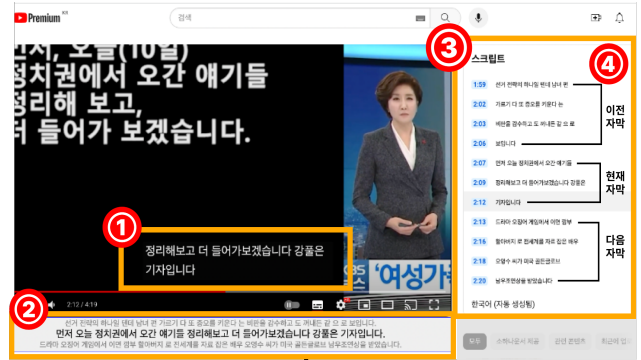


그림 1 서비스 아키텍처

본 연구에서 제안하는 유튜브 자동 자막 제공 서비스의 구조는 그림 1 과 같다. 사용자는 크롬(Chrome) 브라우저에 접속하여 서비스 플러그인을 다운받는다. 이후 시청하고자 하는 유튜브 영상 링크에 접속하여 자막 생성 버튼을 클릭하면 URL 정보가 Flask 서버에 전달되어 자동 자막을

불러온다. 불러온 자막은 mecab 라이브러리를 활용하여 문장으로 재구성하고, CNN 기반의 감정 분석을 진행한다. JSON 형식으로 시간과 자막을 매칭하여 자막 리스트를 만들고 이를 클라이언트에 전달하여 사용자에게 자막을 제공한다. 이때 클라이언트에서는 영상의 현재 시간과 자막 시간을 비교하여 이전, 현재, 다음 자막을 동시에 보여준다.

### 3.1 자막배치 변경



선거 전략의 하나인 텐데 남녀 편 가르기 또 중요키운다는 비판을 감수하고도 깨트린 것으로 보입니다. 먼저 오늘 정치권에서 오간 얘기를 정리해보고 더 들어가보겠습니다 강풀은 기자입니다. 드라마 오징어 게임에서 이런 캠프 할아버지 로 전세계를 사로잡은 배우 오영수 씨가 미국 골든글로브 남우조연상을 받았습니다.

그림 2 서비스 사용 화면

자막이 영상을 가리는 문제를 해결하기 위해 영상 아래로 자막을 배치했다. 그림 2 의 1 번과 2 번을 비교해 보면 1 번은 기존 유튜브에서 제공하는 자동 자막 서비스이고 2 번은 본 서비스에서 제공하는 자막이다. 기존 자동 자막은 배치를 수정할 수 있지만, 영상 외부로 배치할 수 없어서 영상 시청에 불편함이 있다. 또한 자막이 길어지게 되면 영상이 가려지고, 정형화되지 않은 자막창 크기로 인해 영상 시청에 큰 불편이 있다. 그래서 DOM 을 조작하여 영상 밑에 자막 영역을 배치하고, 영상 시청 시간과 자막 리스트를 비교하여 실시간으로 자막을 보여주는 기능을 도입했다. 이로 인해 사용자는 자막이 영상을 가리는 일 없이 시청할 수 있게 된다.

### 3.2 자막 문장화 및 이전과 다음 자막 제공

자막을 문장으로 재구성하고, 이전과 다음 자막을 함께 제공하여 가독성 경험을 높였다. 그림 2 에서 1 번의 자막을 보면 자막이 문장으로 제공되지 않고 무작위로 보여지는 것을 확인할 수 있다. 이는 자막을 보는 사람들로 하여금 가독성을 감소시키는 문제가 있다. 이를 해결하기 위해 Python 의 mecab 라이브러리를 활용하여 그림 2 의 2 번과 같이 문장으로 재구성하였다.

그림 2 의 3 번을 보면 유튜브에서 제공하는 자동 자막 스크립트를 확인할 수 있는데, 이를 순서대로 탐색하며 한글자씩 배열에 추가하고, EF(종결어미) 혹은 EC(연결어미)로 끝나는 문장이 만들어지면 자막을 제공하는 방식으로 구현했다. 최종적으로 그림 2 의 4 번과 같이 이전과 현재, 다음 자막을 구분하고 현재 영상 시청 시간을 기준으로, 3 문장을 함께 제공하였다. 이로 인해 사용자는 문장화 되어진 자막을 볼 수 있고, 이전과 다음 자막을 통해 자막의 문맥을 파악하는 데 도움이 된다.

### 3.3 자막 커스터마이징 및 감정분석

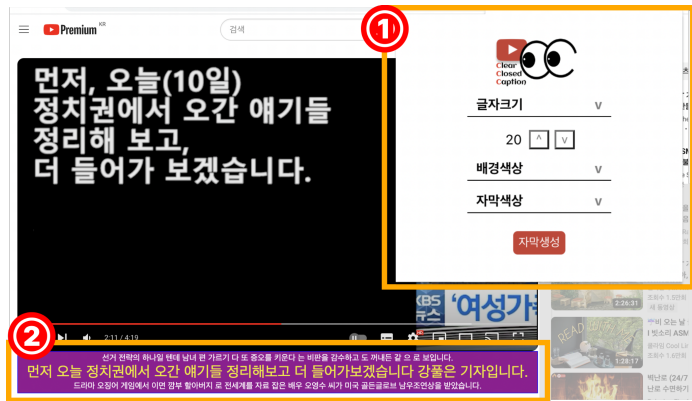


그림 3 커스터마이징 기능

자막의 글자 크기, 배경 색상, 자막 색상을 변경할 수 있는 기능을 제공하여 가독성을 높였고, 감정 분석에 따른 스타일 변화로 동적인 자막 시청 경험을 제공했다. 그림 3의 1 번을 보면 스타일링 할 수 있는 기능들이 제공된다. 사용자는 자신만의 자막 스타일을 커스터마이징 하여서 가독성을 높일 수 있다. 변경 내용은 실시간으로 적용이 될 수 있도록 웹의 Document 객체를 사용했고 DOM의 Style을 비동기 처리로 변경하여 자막 스타일링 기능을 구현했다.

또한 감정 분석을 통해 도출된 긍정 및 부정에 따라 변경되는 자막 색상을 경험할 수 있다. 긍정이 인식되면 노란색으로 보여지고, 부정이 인식되면 파란색으로 보여진다. 또한 Konlpy를 통한 형태소 분석과 CNN Classifier 기법을 활용하여서 감정을 2 가지로 분석했다. 데이터 셋은 네이버 영화 리뷰 데이터를 활용했다. 그림 3의 2 번을 보면 자막 스타일이 노란색으로 변경된 것을 확인할 수 있는데 긍정과 부정의 감정이 각각 70% 이상 느껴지면 스타일을 변경하였다. 긍정일 때는 노란색, 부정일 때는 빨간색을 적용하여서 감정을 표현했다.

## 4. 사용성 평가

### 4.1 사용성 평가 방법

서비스의 사용성을 평가하기 위해 5 명의 인터뷰이를 학교 커뮤니티를 활용하여 모집했고, 사용성 평가는 서비스 다운로드부터 설문조사까지 약 20 분이 소요되었다. 우선 사용성 평가 이전에 서비스 사용방법을 정리해둔 문서를 활용하여 약 10 분간 Zoom 비대면 화상 프로그램에서 사용법을 설명했다. 이후 본격적인 서비스 사용 이전에 유튜브에서 제공하는 자동 자막을 켜 채로 약 5 분 길이의 영상을 먼저 시청하게 한 후 본 연구에서 제시하는 서비스를 사용해 재시청 하도록 하였다.

서비스 사용 이후에 Google Form을 활용하여 설문조사를 진행했다. 설문은 2 가지의 섹션으로 나누었고, 섹션 1은 System Usability Scale (SUS)[8]의 10 가지 항목으로 구성해 전반적인 서비스 사용 만족도를 평가했다. 섹션 2는 서비스의 기능별 만족도를 평가하는 6 가지 항목으로 총 16 가지의 설문 항목으로 구성했다.

### 4.2. 사용성 평가 결과

사용자 평가 점수는 SUS 평가 항목으로 이루어진 섹션 1의 질문들에 대한 결과 값으로만 산정했다. 홀수 항목의 점수 총합에서 5를 빼 값을 X로 두고, 짝수 항목의 점수 총합을 25에서 빼 값을 Y로 두었다. 그 이후  $(X+Y)*2.5$  식을 활용하여 점수를 산정했고, 평균 값을 도출한 결과 70.5로 평균 이상이라는 사용성 평가 결과를 얻을 수 있었다 [8]. 섹션 2 서비스 기능별 만족도 평가에서, 이전과 다음 자막을 함께 제공한 것이 자막의 문맥을 파악하는 데 도움이 되었다는 질문에서 가장 높은 점수를 받았다. 반대로 배경 색상을 변경하는 기능이 가독성을 높이는 데 도움이 되었다는 질문에서는 가장 낮은 점수를 받은 것을 확인할 수 있었다.

## 5. 결론 및 향후 연구

본 연구에서는 동영상 자동 자막 사용 경험을 높이기 위한 서비스를 개발했다. 영상 아래에 자막 영역을 재배치하여 자막이 영상을 가리는 문제를 해결했으며, 자연어 처리 기술을 활용하여 문장으로 자막을 재구성해서 자막을 활용한 문맥 파악이 용이하게 했다. 더불어 자막 스타일 커스터마이징 기능과 감정 분석에 따른 스타일을 적용해 사용자 니즈에 따른 생동감 있는 자막 경험을 제공했다. 사용성 평가를 통해 아이디어의 사용성을 검증했으며, 이를 통해 자동 자막의 기능을 개선하고 시청 경험을 향상시킬 수 있는 중요한 가능성을 제시하였다. 향후 연구로, 보다 정확한 형태소 분석으로 오타 없는 완전한 문장 형태로 자막을 제공하고, 더 풍부한 감정 분석으로 생동감 있는 자막을 제공할 수 있을 것이다. 더불어 새로 구성한 자막의 스크립트를 한 칸에 함께 제공한다면 보다 보기 쉬운 자동 자막 서비스를 구현할 수 있을 것으로 기대된다.

## 6. 참고문헌

- [1] 김선미, 이상원, and 손현정, "동영상 OTT 플랫폼 사업 전략과 정책 방향 : 전문가 델파이 분석," 방송통신연구, pp. 43-73, 2023.
- [2] 박미희, 영상콘텐츠의 자막 구성 전략, 우리어문연구, 75(0), 455-491, 10.15711/WR.75.0.16, 2023.
- [3] 최원희, "2022년 9월 '유튜브' 사용자 4183만, 대한민국 81% 사용", Platum, <https://platum.kr/archives/194239>, 2022.
- [4] 김효진, 박수현, and 임순범, "동영상 대사 자막을 발화자 주변에 자동 배치하는 시스템의 개발," 멀티미디어학회논문지, Vol. 26, No. 6, pp. 732-741, 2023.
- [5] Language Reactor. "Language Reactor". <https://www.languagereactor.com>.
- [6] 김상훈, 민준기, 이해은, and 구명완, "한국어 음성특징을 활용한 감정 인식 기반 자막 생성 솔루션," 한국정보과학회 학술발표논문집, pp. 1877-1879, 개최지, 2023.
- [7] 지영서, 이하람, and 임순범, "음악 감정 분석을 통한 키네틱 타이포그래피 자막 자동 생성 서비스," 멀티미디어학회논문지, Vol. 24, No. 8, pp. 1184-1191, 2021.
- [8] Brooke, John. "SUS: A 'Quick and Dirty' Usability." Usability Evaluation in Industry. 189(3). 189-194. 1996.