

시각장애인을 위한 KoBART 기반의 실시간 뉴스 요약 서비스 시스템 구현

이상록*

Implementation of Real-Time News Summary Service System Based on KoBART for the Visually Impaired

Sangrok Lee*

요약

본 논문은 시각장애인의 뉴스 접근성을 향상하기 위한 한국어 뉴스 요약 서비스 시스템을 구현하였다. 뉴스 요약 서비스는 주기적으로 뉴스 웹 페이지를 접속하여 최신 뉴스들을 검색한 후, 크롤링을 통해 뉴스 원문 기사를 요약 모델로 전달한다. 구현된 KoBART 기반의 뉴스 요약 모델은 AI 허브에서 제공하는 뉴스 데이터셋을 활용하여 미세조정하였다. 그리고, 토큰 크기가 1024 이상인 뉴스 요약을 위해 문단 분할 방식으로 접근하였고, 문단 분할로 인한 문맥 정보 손실을 최소화하기 위해 문단 분할 과정에서 10~20%의 문장을 오버랩하였다. 제인된 시스템은 문단 분할 단위를 512의 토큰 크기로 적용하였고, 문단간에 3개의 문장을 오버랩하는 방식으로 구현하였다. 결과적으로 구현된 시스템으로 시각장애인의 모바일 기기를 활용한 뉴스 요약 서비스가 가능함을 확인하였다.

Key Words : News summary, the Visually Impaired, KoBART, Paragraph Segmentation, Overlap

ABSTRACT

This paper implemented a Korean news summary service system that can improve news accessibility of the visually impaired. The news summary server pe-

riodically accesses news web pages to search for the latest news, and then crawls the original news articles to deliver them to the AI summary model. The implemented KoBART-based news summary model was fine-tuned using the news dataset provided by the AI Hub site. In addition, for news summary with a token size of 1024 or more, we approached the original news using a paragraph division method, and overlapped 10~20% of sentences during the paragraph division process to minimize the loss of contextual information due to paragraph division. The proposed system applied a paragraph division unit of 512 size and was implemented in a way that 3 sentences overlap between paragraphs. As a result, we confirmed that the implemented system enables the visually impaired people to use a news summary service using a mobile device.

I. 서 론

보건복지부에서 작성한 ‘2023년 장애인 실태조사’ 보고서에 따르면 중복장애를 포함한 시각장애인은 27.8 만명으로 추정하고 있고, 지체장애와 청각장애 다음으로 장애인 수가 많은 유형이다^[1]. 또한, 점자 해독이 가능한 시각장애인은 3.3%에 불과하여 점자에 대한 필요성은 줄어들고, 스마트 폰 또는 태블릿 등의 모바일 기기를 95.2%가 이용하므로 대부분 시각장애인이 정보통신기기를 사용하는 것으로 나타났다. 따라서 최근에는 시각장애인을 위해서는 점자 방식이 아닌 모바일 기기 기반의 다양한 서비스 개발이 필요한 상황이다.

2015년부터 2024년까지 10년간 시각장애인을 대상으로 한 AI 기술 개발 및 활용 관련 국내외 연구 동향을 분석한 결과 국내 연구는 30개에 불과한 것으로 나타났다^[2]. 주요 연구 분야는 머신 러닝과 컴퓨터 비전 등을 적용하여 시각 정보를 처리하는 대화형 솔루션 개발로 나타났다. 논문 [3]에서는 시각장애인의 활동성을 보장하기 위한 보조 시스템을 임베디드 시스템과 딥러닝 기반의 학습 모델을 활용하여 구현하였다. 깊이 측정용 카메라를 이용하여 주변 환경과 물체들을 인식하고, 마이크와 스피커를 통해 인터페이스함으로써 시각장애인의 보조시스템으로 활용 가능성을 확인하였다.

* 본 논문은 2024년도 신한대학교 학술연구비 지원으로 연구되었음.

◆ First Author : (ORCID:0000-0001-7016-2572) Shinhan University, Department of Electronic Engineering, srlee@shinhan.ac.kr,
부교수, 종신회원

논문번호 : 202506-130-E-LU, Received June 4, 2025; Revised June 25, 2025; Accepted June 25, 2025

본 논문에서는 시각장애인이 모바일 정보통신기기 를 사용하는 환경에서 정보 접근성 향상을 위해 한국어 기반의 자연어 처리 모델인 KoBART 모델을 적용하여 실시간 뉴스 요약 서비스 시스템을 구현한다.

II. 시스템 설계

그림 1은 시각장애인에게 실시간 뉴스 요약 서비스를 위한 시스템 구성이고, 크게 시각장애인이 사용하는 모바일 기기와 NLP 서버로 구성된다. 서비스 시스템의 모바일 기기에는 서비스 앱이 설치되고, NLP 서버는 뉴스 요약 모델과 웹 서비스 기능으로 구성된다. 뉴스 요약 서비스 시스템의 동작 과정은 다음과 같다. 먼저, 서버에서 뉴스 요약 모듈은 뉴스 웹 사이트를 주기적으로 크롤링하여 뉴스의 기본 정보와 요약문을 생성한다. 주기적으로 생성된 정보들은 API를 통해 Flask 서버로 전송하고, 사용자가 사용할 수 있도록 모바일 기기의 앱으로 전달된다. 모바일 기기에 설치된 시각장애인 전용 앱은 업로드된 정보들을 활용하여 뉴스 제목을 노출 한다. 시각장애인이 전용 앱을 실행하여 뉴스를 선택하면, 뉴스 요약문을 TTS (Text-To-Speech)로 읽어 준다.

그림 2는 NLP 서버에 구현된 뉴스 요약 모델의 흐름도를 보이고 있다. 먼저, 뉴스 홈페이지를 주기적으로 크롤링한다. 불필요한 광고를 제외하고, 새로운 뉴스가 업데이트되면 뉴스의 URL을 추출한 후, HTML 파싱을 통해 뉴스 기사의 정보들을 수집하여 데이터베이스에 저장한다. 데이터베이스는 언론사 번호, 뉴스 기사 번호, 기사 제목, 날짜 및 시간, 기자 정보, 카테고리 등의 기사 정보와 기사 원문으로 구성된다. 그리고, 기사 원문은 한국어 뉴스 요약을 위한 KoBART 기반의 모델을 통해 요약문을 생성하고, 요약문도 서비스를 위해 데이터베이스에 저장한다. 마지막으로 데이터베이스에 저장된 모든 데이터는 JSON 형식으로 REST API를 통해 전달한다.

한국어 뉴스 요약 모델은 허깅 페이스에서 배포된

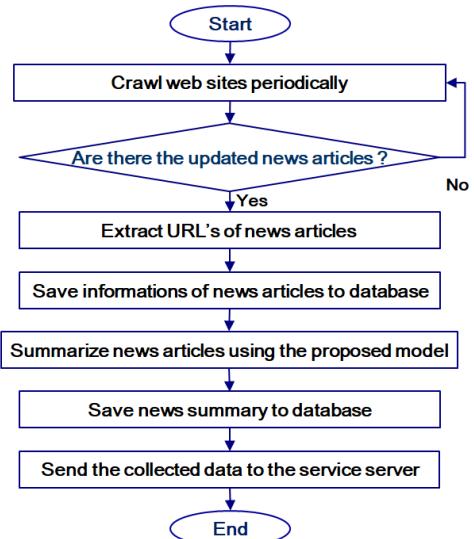


그림 2. 뉴스 요약 서버 흐름도

Fig. 2. Flowchart of news summary server

KoBART 모델과 토크나이저를 기반으로 구현하였다^[4]. 또한, 요약 성능 향상을 위해 기존 KoBART 모델은 AI 허브에서 제공하는 학습용 데이터셋에서 신문 기사와 관련된 데이터셋을 적용하여 미세 조정하였다. 그리고, 기존 KoBART 모델이 토큰 길이가 1024 이상의 장문 뉴스는 요약이 어려운 문제점을 개선하기 위해 원문을 분할하여 요약하는 방식으로 접근하였다^[5]. 본 논문에서 구현된 뉴스 요약 모델은 뉴스 원문의 토큰 길이가 1024 이하인 경우에는 한 번에 요약문을 생성하고, 그 이상의 장문 뉴스는 일정 크기의 문단으로 분할하여 다단계로 요약 모델을 적용하여 요약문을 생성하였다. 논문 [5]의 연구 결과에서 보듯이 장문 뉴스 처리 과정에서 원문 뉴스를 분할할 때 문맥 정보의 연결성이 손상되어 루지 스코어의 성능이 저하되는 문제점이 발생하였다. 따라서 본 논문에서는 분할로 인한 문맥 정보의 손실을 최소화하기 위해 분할되는 문단 간에 일부 문장을 오버랩(overlap)하는 방법을 적용하였고, 뉴스 원문 대비 오버랩 문장의 비율은 실험을 통해 확인하였다.

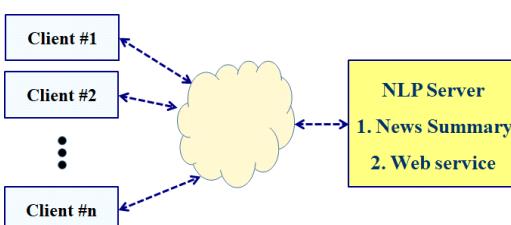


그림 1. 뉴스 요약 서비스 시스템 구성

Fig. 1. Configuration of news summary service system

III. 시스템 구현 및 검증

시각장애인을 위해 제인된 뉴스 요약 서비스 시스템의 구현 및 성능 평가에 사용된 시스템 사양은 표 1과 같고, GPU는 NVIDIA GeForce RTX™ 4070 Ti를 사용하였다. 또한, 시각장애인을 위한 뉴스 서비스는 네이버 포털에 올라오는 뉴스를 대상으로 진행하였고, 서비스할 뉴스의 카테고리는 정치, 사회, 경제, 세계, IT/과

표 1. 뉴스 요약 서비스 시스템 사양
Table 1. Specification of news summary service system

Component	Specification
CPU	AMD Ryzen™ 7 7700 8-Core
GPU	NVIDIA GeForce RTX™ 4070 Ti
Memory	DDR5 64G PC5-44800
OS	Ubuntu 22.04.05 LTS
Software	PyTorch, Python, Andoriod studio, BeautifulSoup, SQLite3

학, 생활/문화, 오피니언 분야를 포함한다.

그림 3은 구현된 KoBART 기반의 장문 뉴스 요약 모델의 성능을 분석하기 위해 뉴스 원문의 분할 크기가 {128, 256, 512}이고, 문단에서 오버랩 적용 유무에 따른 ROUGE-L 스코어를 측정한 결과이다. 단, 뉴스 원문 크기에 대한 오버랩 문장의 비율은 뉴스 원문의 분할 크기에 따라 10%에서 20% 이내가 되도록 설정하였다. 결과적으로 뉴스 분할 크기가 128일 때는 1문장, 256일 때는 2문장, 그리고 512일 때는 3문장을 오버랩하였고, 실험에 적용된 데이터셋에서 비율로 환산하면 각각 16.95%, 16.65%, 11.36%이다.

문단 분할 크기가 {128, 256, 512}일 때 오버랩을 적용하면 ROUGE-L 스코어가 각각 0.27%, 1.51%, 0.39% 향상됨을 확인하였다. ROUGE-L 스코어 측면에서는 분할 크기를 256으로 할 때 성능 향상이 가장 높지만, 분할 크기가 작아질수록 요약 소요 시간이 비례적으로 증가한다. 그림 3에서 보듯이 오버랩을 적용한 256분할에서 ROUGE-L 스코어가 34.61로 가장 높지만, 구현한 시스템에서는 오버랩을 적용한 512 분할을 적용하였다. 이는 ROUGE-L 스코어는 256 분할에 비해 0.23% 낮지만, 요약 소요 시간은 분할 비율에 거의 비례하여 낮아지기 때문이다.

그림 4는 구현된 시스템의 실행 과정에서 서버에 저

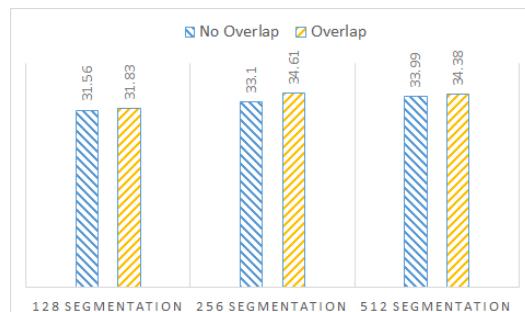


그림 3. 오버랩과 분할 비율에 따른 루지 스코어 비교
Fig. 3. Comparison of ROUGE-L score according to overlap and segmentation ratio

category	paper	article_title	article_text	translated_res	author	article_date	article_time	vif_container
Search column	Search column	Search column	Search column	Search column	Search column	Search column	Search column	Search column
1. 서론	매일신문	TK 행정부는 논의 중... 해원도 강제	부동산 특별연락... 여러분은 대체로 사랑해...	임금기 기자 younki@... 여전히 청와대에서 이...	여전히 청와대에서 이...	2024-05-20	16:41:06	http://newsnaver...
2. 세부	매일신문	여전히 청와대에서 이...	여전히 청와대에서 이...	여전히 청와대에서 이...	여전히 청와대에서 이...	2024-05-20	16:42:01	http://newsnaver...
3. 자료	매일신문	‘경찰의 소는 늘 그렇지... 전경관은 이미 그녀였...	경찰관은 이미 그녀였...	경찰관은 이미 그녀였...	경찰관은 이미 그녀였...	2024-05-20	16:43:06	http://newsnaver...
4. 서론	매일신문	본문은 다음으로 정리... 본문은 다음으로 정리...	본문은 다음으로 정리...	본문은 다음으로 정리...	본문은 다음으로 정리...	2024-05-20	16:44:01	http://newsnaver...
5. 자료	매일신문	부동산 특별연락... 부동산 특별연락...	부동산 특별연락...	부동산 특별연락...	부동산 특별연락...	2024-05-20	16:45:06	http://newsnaver...
6. 경계	매일신문	여전히 청와대에서 이...	여전히 청와대에서 이...	여전히 청와대에서 이...	여전히 청와대에서 이...	2024-05-20	16:46:01	http://newsnaver...
7. 경계	국민일보	SBS 대화기 2종 출판... 여전히 청와대 GBC 조작...	여전히 청와대 GBC 조작...	여전히 청와대 GBC 조작...	여전히 청와대 GBC 조작...	2024-05-20	16:46:01	http://newsnaver...
8. 트랙	전자신문	전자신문 제100주... 개인 정보는 개인 정보...	개인 정보는 개인 정보...	개인 정보는 개인 정보...	개인 정보는 개인 정보...	2024-05-20	16:47:01	http://newsnaver...
9. 자료	뉴스1	인터넷광고대행사... [창립 1주년 기념]	[창립 1주년 기념]	[창립 1주년 기념]	[창립 1주년 기념]	2024-05-20	16:47:35	http://newsnaver...
10. 세부	오마이뉴스	‘조각당’ 대통령 비서... [아이 대행장, 팔...]	[아이 대행장, 팔...]	[아이 대행장, 팔...]	[아이 대행장, 팔...]	2024-05-20	16:48:01	http://newsnaver...
11. 정리	조선일보	정부, 7월 14일 폭격... 정부는 광복절 차운 시...	정부는 광복절 차운 시...	정부는 광복절 차운 시...	정부는 광복절 차운 시...	2024-05-20	16:49:01	http://newsnaver...
12. 서론	매일경제	한국개발원은 최근... [서울=뉴시스] 김승관...	[서울=뉴시스] 김승관...	[서울=뉴시스] 김승관...	[서울=뉴시스] 김승관...	2024-05-20	16:49:32	http://newsnaver...
13. 세부	매일경제	국내외 기관들은... [국내외 기관들은...]	[국내외 기관들은...]	[국내외 기관들은...]	[국내외 기관들은...]	2024-05-20	16:50:01	http://newsnaver...
14. 경계	마이크로소프트	관련 기관은 전문... [관련 기관은 전문...]	[관련 기관은 전문...]	[관련 기관은 전문...]	[관련 기관은 전문...]	2024-05-20	16:50:05	http://newsnaver...
15. 자료	조선일보	조선일보 대체로 ... [백신 아동방광증후군...]	[백신 아동방광증후군...]	[백신 아동방광증후군...]	[백신 아동방광증후군...]	2024-05-20	16:51:41	http://newsnaver...
16. 결론	파이낸셜뉴스	S/S 40 한글 폰트... [22개 한글 폰트...]	[22개 한글 폰트...]	[22개 한글 폰트...]	[22개 한글 폰트...]	2024-05-20	16:52:03	http://newsnaver...

그림 4. 뉴스 요약 데이터베이스 예시 화면

Fig. 4. Example screen of news summary database

Article Titles

- [부산시와 광주시, 소·부·장 관역협력 손잡았다](#)
- [기아 오프로드 광주 지역 섬광세대 육성 위한 후원금 전달](#)
- [박찬대 “해병대 특별경찰 제의결에 힘드니 동참해달라”](#)
- [SK 지난해 사회적 가치 16.8조 창출...사회 성과·분야 역대 최대](#)
- [‘의정부 호원초 교사 사망’ 학부모·학교 관계자 모두 무혐의](#)
- [용인중·고시장 창설 약속履行 청취자 10팀 모집...27일까지 신청](#)
- [CU, ‘서울세계도시문화축제’ 참여...“K편의점 문화 전파”](#)
- [\[기자회견\] “안 멈추고 계속 자랐다”...제주 마늘 수확 현장 가보니](#)

그림 5. 모바일 기기의 예시 화면

Fig. 5. Example screen of mobile device's app

장된 데이터베이스의 예시 화면이다. 뉴스 기사에 대한 기사 정보, 뉴스 원문, 뉴스 요약문이 데이터베이스에 저장된 것을 확인할 수 있다. 그림 5는 시각장애인 모바일 기기에서 사용하는 전용 앱의 예시 화면을 보이고 있다. 최신 뉴스들의 제목이 나열된 것을 볼 수 있고, 시각장애인이 뉴스 제목을 터치하면 TTS로 뉴스 제목을 읽어준다. 동일한 뉴스 제목을 다시 한번 터치하면 뉴스 요약문을 읽어주는 방식으로 서비스된다.

IV. 결 론

정보 홍수의 시대에 시각적 인터페이스 조작이 어려운 시각장애인에게 뉴스 접근성을 향상시킬 수 있는 한국어 뉴스 요약 서비스 시스템을 구현하였다. 뉴스 요약 서비스는 주기적으로 뉴스 웹 페이지를 접속하여 최신 뉴스들을 검색한 후, 크롤링을 통해 뉴스 원문 기사를 요약 모델로 전달한다. 구현된 KoBART 기반의 뉴스 요약 모델은 AI 허브에서 제공하는 뉴스 데이터셋을 활용하여 미세조정하였다. 그리고 토큰 크기가 1024 이상인 뉴스의 요약을 위해 문단 분할 방식으로 접근하였고, 문단 분할로 인한 문맥 정보 손실을 최소화하기 위해 문단 분할 과정에서 10~20%의 문장을 오버랩하였다. 제안된 시스템은 문단 분할 단위를 512로 적용하였고, 문단간에 3개의 문장을 오버랩하는 방식으로 구현하였다. 결과적으로 뉴스 접근성이 낮은 시각장애인들이 구현된 시스템을 활용함으로써 사회적 약자들의 불평등 해소에 긍정적 영향을 줄 것으로 기대된다.

References

- [1] Ministry of Health and Welfare, *2023 Survey on the Status of Persons with Disabilities*, 2023.
- [2] J. Park and Y. Lee, “Applications of artificial intelligence technology for individuals with visual impairments: A literature review using text mining techniques,” *The Korean J. Visual Impairment*, vol. 40, no. 3, pp. 1-19, 2024.
- [3] H.-J. Yang, S.-W. Jang, J.-S. Park, J. Kim, and J. Kwon, “A study on the development of real-time embedded assistance system based on deep learning for the blind,” *J. Rehabilitation Welfare Eng. & Assistive Technol.*, vol. 16, no. 1, pp. 27-36, Feb. 2022.
- [4] Hugging Face, <https://huggingface.co/gogamza/kobart-base-v1>, 2024.07.12.
- [5] M.-K. Kim and S. Lee, “Implementation of KoBART-based real-time long news summarization system using text segmentation,” *J. Ind. Technol. Res.*, vol. 29, no. 3, pp. 27-35, Sep. 2024.