

추출요약 기법을 사용한 뉴스 기사 요약 애플리케이션의 설계 및 구현

문봉관*, 임한규**, 박은주***

Design and Implementation of News Article Summarization Application using Extraction Summarization Techniques

Bongkwan Moon*, Hankyu Lim**, and Eunju Park***

본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2019-0-01113)

요약

정보를 제공함에 있어서는 정확한 정보를 신속하게 제공해야 한다. 본 논문에서는 추출요약 기법을 이용하여 문서의 전체 내용을 빠르게 파악할 수 있는 문서 요약 애플리케이션을 설계하고 구현하였다. 본 논문의 문서 요약 애플리케이션은 웹 크롤링을 사용하여 관심 있는 뉴스 기사의 제목과 본문을 수집한다. API를 사용하여 수집된 내용을 파이썬 Flask 서버로 넘겨주어 추출 요약을 실행하고, 요약된 내용은 애플리케이션으로 전송하여 화면에 출력함으로써 사용자는 요약된 기사의 내용을 확인함으로써 전체 기사를 읽을 때보다 빠르게 기사 내용의 확인이 가능하다. 또한, 새로운 기사 업로드시 사용자가 즉시 인지하도록 알림 기능을 추가와 원하는 경우에만 요약 기능을 확인할 수 있는 기능, 팝업 창 기능 등을 추가하여 사용자 편의성을 높였다. TextRank 알고리즘을 이용한 본 시스템은 정보의 신속성과 정확성을 동시에 충족시킬 수 있는 장점이 있다.

Abstract

When providing information, it is necessary to deliver accurate information promptly. In this paper, we designed and implemented a document summary application that can quickly grasp the entire contents of a document using the abstraction summary technique. The document summary application of this paper uses web crawling to collect the titles and body of news articles of interest. Using the API, the collected content is passed to the Python Flask server to run the abstraction summary, and the summarized content is sent to the application and displayed on the screen. It is possible to check the contents faster than reading it directly. Additionally, to enhance user convenience, features such as notifications to alert users immediately when a new article is uploaded, an option to view the summary only when desired, and pop-up windows were added. The system utilizing the TextRank algorithm has the advantage of simultaneously satisfying the speed and accuracy of information delivery.

Keywords

document summary, extract summary, API, Flask, web crawling, application notification

* 국립안동대학교 멀티미디어공학과 석사
- ORCID: <https://orcid.org/0009-0005-1235-7933>
** 국립안동대학교 소프트웨어융합학과 교수
- ORCID: <https://orcid.org/0000-0003-1755-4651>
*** 국립안동대학교 SW융합교육원 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-4579-5535>

· Received: Apr. 05, 2024, Revised: May 20, 2024, Accepted: May 23, 2024
· Corresponding Author: Eunju Park
SW Convergence Education Center, Andong National University, Korea
Tel.: +82-54-820-6307, Email: eunju@anu.ac.kr

I. 서 론

4차 산업혁명 시대를 맞아 실시간으로 생성되는 무수히 많은 정보 속에서 사람들은 각자 자기만의 방식으로 필요한 정보를 얻는다[1]. 2020년 한국언론진흥재단의 ‘언론수용자 조사’에 따르면 20대에서 40대의 사람들은 뉴스 정보를 얻기 위한 수단으로 스마트폰을 가장 많이 사용하고, 하루 평균 인터넷 사용시간은 95.3분을 사용하며, 뉴스 검색 평균시간은 21.6분이며 원하는 뉴스 정보는 신속하게 얻고자 하는 것으로 조사되었다[2]. 이에 현대사회에서는 빠르고 정확하게 많은 정보를 획득하고 제공하는 방법에 대한 연구의 필요성이 대두되고 있으며, 이런 연구 가운데 하나가 요약이다[3].

요약은 내용의 핵심을 간추려서 제공하는 것으로 크게 생성요약과 추출요약 방법이 있다. 생성요약은 문서를 요약할 때 새로운 자연어를 만들어내는 방법으로 아직까지는 인간 수준의 언어 구사 능력을 가질 수 있게 하는 자연어 처리 연구가 완성되지 않아 한계를 가지는 요약 방법이다[4].

추출요약은 단일 문서 혹은 다중 문서에서 핵심이 되는 문장을 추출해서 요약하는 것으로 안정적인 결과를 가질 수 있어 생성요약 방법보다 많이 사용된다[5]. 추출요약의 TextRank 알고리즘은 대표적인 추출요약 방법 중 하나로 문장을 그래프로 만들어 가중치를 매기는 랭킹 알고리즘을 사용하는 요약 방식이다[6]. TextRank 알고리즘은 주어진 문서 집합 안에서 단어들이나 문장을 선택함으로써 요약된 결과가 원래 문장이 아닌 것은 가지지 않는다는 장점을 가진다. 이에 본 논문에서는 추출요약 방법의 TextRank 알고리즘을 이용하여 요약 기능을 구현하고, 파이썬 서버와 안드로이드 스튜디오를 연결하여 데이터를 API로 전송함으로써 사용자가 스마트폰으로 뉴스 기사를 핵심 내용을 3줄로 요약한 요약문을 제공받을 수 있는 시스템을 설계하고 구현하였다. 또한 알림 설정 기능을 추가하여 사용자가 직접 애플리케이션에 접속하지 않아도 새로운 뉴스 기사를 신속하게 접할 수 있게 하였다. 본 논문의 시스템의 구현으로 사용자는 기존에 뉴스 기사에 대한 정보를 얻기 위해 애플리케이션에 접속하여

뉴스 기사를 검색하는 시간을 절약할 수 있으며, 뉴스 기사 전체를 읽지 않아도 추출 요약을 통해 내용의 핵심을 전달받아 내용을 빠르게 파악하는 것이 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 구현할 시스템에 관한 관련 연구를 서술하였으며, 3장에서는 시스템에 대한 설계, 4장에서는 시스템의 구현에 대하여 기술하였다. 5장은 결론으로 본 논문에서 구현한 시스템에 대한 시사점에 대하여 기술하였다.

II. 관련 연구

2.1 자연어 처리

자연어란 인간의 의사소통을 위한 기호의 전달 체계로 음성이 기호화를 통해 의미를 전달하는 두뇌 기관의 작동을 컴퓨터가 모사하도록 하는 기술이 자연어 처리이다[7]. 자연어 처리는 많은 양의 자료들이 기반이 되어야 하며, 자료 사이의 규칙을 찾고 이런 규칙을 기반으로 새롭게 문장이나 글을 구성하는 등 여러 작업을 할 수 있다.

본 논문에서 사용하는 TextRank 알고리즘은 가장 기본적인 자연어 처리의 방법 중 하나이다. TextRank 알고리즘은 전처리과정에서 조사와 동사를 제거한 명사 형태의 데이터들로만 구성된 후, 데이터마다 가중치를 부여하여 원하는 데이터를 추출하는 방식의 간단한 머신러닝 작업이다.

2.2 문서요약

문서요약은 문서를 용도에 맞게 알맞은 형태로 축약해서 제공하는 것을 의미하며, 자연어 처리의 한 분야로 추출요약과 생성요약의 두 가지 방법을 많이 사용한다[8][9].

추출요약은 원본 문서에서 제목, 단어의 출현 빈도, 문장 간의 유사도, 그래프 등을 이용하여 주요 핵심 문장 몇 개를 추출하는 방식으로, PageRank 알고리즘과 TextRank 알고리즘 등을 사용한다[10][11]. 추출요약은 문서 내에 존재하는 문장에 가중치를 다르게 매겨 핵심이 되는 문장을 그대로 추

출해서 제공하기 때문에 문서와 관련이 적은 문장을 추출할 가능성이 적으며, 문장을 직접 변형하지 않기 때문에 본래 문장의 의미가 훼손되지 않는다는 장점이 있다[12].

생성요약은 사람이 직접 문서를 보고 내용을 이해한 다음 요약문을 작성하는 것처럼 문서 내에 존재했던 문장이 아닌 새로운 문장을 사용하여 사용자가 이해하기 쉽도록 내용을 요약해서 제공하는 방법이다[13]. 이 방식은 사람이 문장을 구성하듯이 문법적으로 정확하고 자연스러운 문장을 구성하는 것이 매우 까다롭다. 각 나라의 언어가 가지는 특성을 이해한 다음 문장을 구성해야 표현하고자 하는 내용의 요약이 문서의 내용과 어긋나지 않기 때문이다. 따라서 이 방식은 실제 구현에 상당히 난이도가 있는 편이지만 최근 들어 머신러닝과 딥러닝 등을 이용한 연구가 활발히 진행되고 있다. 입력 시퀀스를 다른 형식의 시퀀스로 출력, 변환하는 모델인 seq2seq 모델과 seq2seq 모델의 단점을 보완하기 위해 문장의 생성에 문서를 한 번 더 참조하도록 고안된 모델인 attention 모델, 복사 방법론, 포인터 네트워크 모델 등 다양한 모델들이 존재한다[14][15].

본 논문에서는 추출요약의 TextRank 알고리즘을 사용하여 문서에서 핵심 문장이 되는 3개의 주요 문장을 추출하고, 사용자에게 제공하기 위하여 API를 사용한 데이터 통신을 통해 안드로이드에 요약된 문장을 제공하도록 설계하였다. TextRank 알고리즘은 각 문장들 간의 유사도와 기사 전체에서 빈도가 높은 단어를 측정해서 핵심 문장을 찾아낸다. 핵심 문장일수록 다른 문장에 반복적으로 사용되는 단어가 여러번 포함된 문장일 가능성이 높다. 이 알고리즘을 사용해서 뉴스 기사를 요약하는 이유는 기사에 핵심이 될 수 있는 단어를 포함하는 문장 자체를 추출하기 때문에 사용자는 기사가 전달하려는 기사 내용의 변형 없이 제공받을 수 있다.

2.3 Open API

Open API(Application Programming Interface)는 인터넷 상의 서비스, 정보, 데이터를 외부의 사용자 및 개발자가 사용할 수 있도록 개방해 놓은 API를

일컫는다[16]. Open API는 보통 웹 서비스의 형태로 배포되며, 대표적인 방식으로는 SOAP API와 REST API 등이 있다.

SOAP API는 서비스를 이용하는 클라이언트와 서비스를 제공하는 서버 쌍방이 SOAP의 생성/해석 엔진을 가지기 때문에 다른 환경에서도 객체의 호출이 가능한 특징을 가지고 있으며, XML 형태의 데이터만을 사용할 수 있다. 반면에 REST API는 SOAP보다 웹상의 리소스를 별도 계층 없이 전송할 수 있으며, XML뿐만 아니라 JSON 등의 다양한 데이터의 사용이 가능하기 때문에 SOAP보다 사용이 간편하다는 장점을 가지고 있다[17]. 이러한 REST API의 장점을 고려하여 본 논문에서는 안드로이드 애플리케이션으로부터 JSON 형식으로 된 데이터를 전달받아 요약을 실행한 뒤, 요약 처리된 데이터를 다시 안드로이드 애플리케이션에 JSON 형식으로 전송하는 REST API를 설계하고 구현하였다.

2.4 Flask

Flask는 파이썬으로 작성된 마이크로 웹 프레임워크로 소스코드를 이해하기 쉽고 사용에 적합한 프레임워크이다[18]. 간편하지만 견고한 웹 프레임워크로 사용할 패키지를 선택해서 코드를 작성해 개발할 수 있어 웹과 파이썬을 동시에 사용하고자 할 때 주로 사용한다[19]. Flask는 파이썬을 이용한 기능 구현과 안드로이드와 연동이 모두 가능한 프레임워크로 본 논문에서는 Flask를 채택하여 안드로이드 스튜디오에서 크롤링한 웹페이지 URL을 Flask 서버로 전송해 파이썬 코드로 추출 요약한다. 결과로 반환된 추출 문장은 다시 Flask에서 안드로이드 스튜디오로 전송하도록 설계하였다.

2.5 웹 크롤링

웹 크롤링은 특정 웹사이트의 웹 문서를 자동으로 수집하는 기술로 웹 데이터 증가에 따라 중요성이 커지고 연구가 활발한 분야이다[20]. 웹 크롤링으로 웹 사이트에서 수집할 내용을 선정하고 수집하고자 하는 부분의 소스코드를 분석하여 파이썬을 이용해 자동으로 수집할 수 있다.

본 논문에서는 크롤링 기술을 활용하여 웹 사이트의 뉴스 기사 URL을 수집하여 해당 URL에 포함되어 있는 뉴스 기사의 제목과 본문은 추출요약에 사용될 수 있도록 변수에 따로 저장한다. 그리고 수집된 뉴스 기사 제목과 본문을 애플리케이션에서 볼 수 있도록 화면에 출력하도록 설계하였다.

2.6 관련분야 연구

텍스트 요약은 꾸준히 연구되고 있으며, 다양한 분야에서 활용되고 있다. 텍스트 요약의 예로 수사 구조를 이용한 텍스트의 요약 시스템과 맥락을 탐색을 통한 요약 시스템에 대한 연구, 자연어 원문 구문 구조에서 구 구조를 축소하여 문서 요약에 적합한 구조를 산출하는 텍스트 문서에서 핵심 정보를 뽑아내는 문서 자동 요약 연구 등이 있다[21][22].

추출 요약에 대한 연구는 키워드 추출과 문장 추출로 나뉜다. 블로그나 SNS 등의 소셜 리뷰는 최근 트렌드를 반영한 신조어나 고유명사가 많기 때문에 사전에 포함되지 않아 기존의 형태소 분석기에서 잘 인지하지 못한다. 따라서 긴 시간이 소요되는데 이러한 단점을 보완하기 위한 응집도 점수를 기반으로 한 소셜 빅데이터의 효율적인 키워드 추출과 같은 키워드 추출에 대한 연구가 있다[23][24]. 또 다른 연구로 위키피디아 데이터를 이용한 한국어-영어 병렬 문장 추출 등과 같은 문장 추출에 대한 연구 등 더 자세하며 세부적인 요구사항들에 대한 연구가 진행되고 있다[25].

TextRank 알고리즘은 키워드 추출과 문장 추출의 기능을 제공하며, 자연어 처리 분야에서 TextRank 알고리즘을 응용한 시스템에 대한 연구가 진행되고 있다[26]. TextRank 알고리즘을 이용한 자동으로 회의록을 생성하는 시스템, 기존의 TextRank 알고리즘을 사용하지만 단순한 단어 빈도로 문장 사이의 가중치를 정하는 방법에서 더 개선된 두 문장 사이의 관련성 정도를 측정하여 개선된 TextRank 모델, 단일 문서 내에서 문장 간의 수반 관계를 고려하며 TextRank를 기반으로 하는 TextRank-NLI 모델 등에 대한 연구들이 진행되었다[27]-[29].

기존의 TextRank 알고리즘을 이용한 개선된 연구

들이 나오고 있으나 본 논문에서는 개선된 모델들이 본 시스템에 적합한지에 대한 조사가 이루어지지 않아 기본적인 TextRank 알고리즘 모델을 사용한 스마트폰 애플리케이션으로 구현하였다.

III. 설 계

본 논문의 뉴스 기사 요약 애플리케이션은 안드로이드에서 웹크롤링을 통해 수집한 문서의 내용을 Json 방식으로 저장하여 파이썬 API를 통해 파이썬 Flask 서버에 전송하여 변수로 저장한다. 이후 해당 문서 내용을 TextRank 알고리즘을 적용시켜 문서 요약을 수행하고, 요약된 결과 문장들을 다시 파이썬 API를 통해 안드로이드에 전달하여 사용자가 볼 수 있도록 화면에 표시하도록 설계하였다.

본 논문에서 구현할 뉴스 기사 요약 시스템의 뉴스 기사를 요약하는 핵심동작의 DFD(Data Flow Diagram)와 흐름도는 그림 1과 같다.

DFD와 흐름도 이외에도 시스템 구현을 위하여 유스케이스 다이어그램, 클래스 다이어그램, UI 설계를 진행하였다.

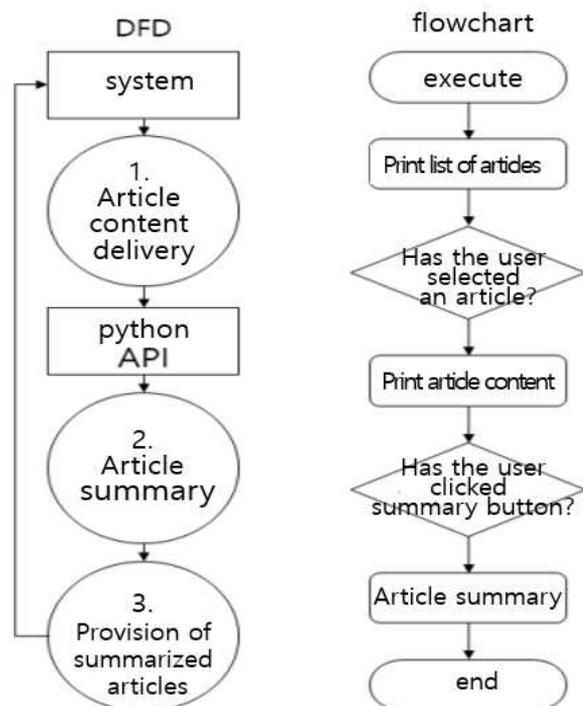


그림 1. 요약 시스템 DFD(좌)와 흐름도(우)
Fig. 1. Summary system DFD and flow chart

3.1 다이어그램 설계

유스케이스 다이어그램은 사용자의 관점으로 시스템과의 상호작용을 설명하기 위한 다이어그램을 설계하는 것으로 시스템에서 제공되는 기능이나 서비스들을 명시한다. 그림 2는 본 논문의 애플리케이션에 대한 유스케이스 다이어그램이다.

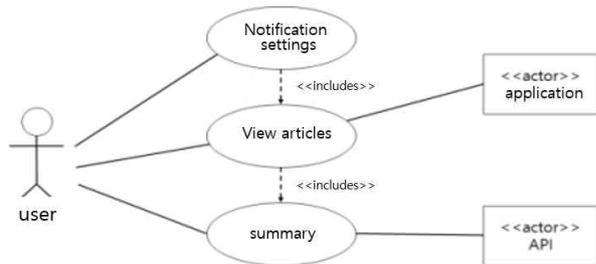


그림 2. 유스케이스 다이어그램
Fig. 2. Usecase diagram

사용자는 스마트폰 애플리케이션에서 뉴스 기사를 볼 수 있고, 알림설정 기능을 통해 기사에 대한 알림을 받아 알림 클릭시 등록된 기사를 확인할 수 있다. 또한 추가적인 기능으로 API에서 요약된 기사를 제공받을 수 있다.

클래스 다이어그램은 시스템에서 사용되는 클래스들을 정리하여 작성한 것으로 시스템의 구조를 한 눈에 파악할 수 있다는 장점을 가진다. 그림 3은 본 논문의 뉴스 요약 애플리케이션의 클래스 다이어그램이다.

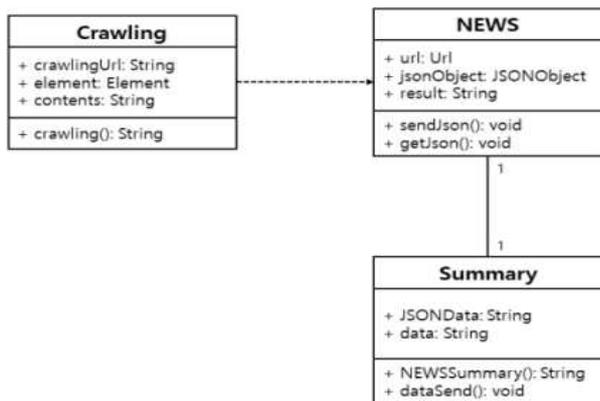


그림 3. 클래스 다이어그램
Fig. 3. Class diagram

본 시스템에서는 기사 내용을 가져오기 위한 Crawling 클래스, 기사 내용을 파이썬 Flask 서버에

전달하기 위한 NEWS 클래스, 기사 내용을 요약하기 위한 Summary 클래스를 가지며 각 클래스에 사용되는 변수 및 함수들이 클래스의 하위 목록에 작성됨을 확인할 수 있다.

3.2 알림 설정

알림은 사용자에게 미리 알림을 보내 다른 사람과 소통할 수 있도록 하며, 애플리케이션에서 보내는 시기적절한 기타 정보를 제공하기 위해 안드로이드가 앱의 UI 외부에 표시하는 메시지이다[30]. 사용자는 알림을 탭하여 앱을 열거나 알림에서 바로 특정 작업을 실행할 수 있다.

기기에 표시되는 알림 형식은 상태표시줄 및 알림 창, 헤드업 알림, 잠금 화면 알림 그리고 앱 아이콘 배지로 총 4가지가 있으며 각각 다른 특징과 역할을 가지고 있다.

상태표시줄 및 알림 창 형식은 스마트폰의 가장 위인 상태 표시줄에 아이콘으로 표시되며 상태 표시줄을 아래로 스와이프 할 경우 해당 알림 창에서 간략하게 요약된 알림을 볼 수 있는 형식이다. 헤드업 알림은 다른 앱보다 우선적으로 화면에 표시하여 사용자가 즉시 알 수 있게 표현하는 것이다. 잠금 화면 알림은 잠금 화면인 상태에서 앱이 게시하는 알림을 표시할 것인지를 설정한 후 잠금 화면인 상태에서 해당 알림을 표시하는 것이며, 앱 아이콘 배지는 사용자가 직접 앱 아이콘을 길게 눌러 앱의 알림을 보는 방식이다.

본 논문의 시스템에서는 사용자가 새로운 뉴스 기사를 직접 파악할 때까지 기다리지 않고 바로바로 기사가 등록된 것을 확인할 수 있도록 상태표시줄 및 알림 창 형식을 도입했고, 이를 통해 사용자는 뉴스 기사를 더욱 신속하게 접해 원하는 정보를 얻을 수 있다.

IV. 구 현

4.1 기능 구현

본 논문의 뉴스 기사 요약 시스템은 다음 순서로 구현하였다.

먼저 기사의 요약을 위해 기사의 링크로 접속하여 데이터들을 크롤링을 통해 가져온다. 그리고 파싱을 통해 크롤링 한 데이터 중 사용하고자 하는 내용만을 Json 형식의 데이터로 추출하고, 파이썬 REST API를 통해 파이썬 Flask 서버에 전달한다. 전달된 내용은 TextRank 알고리즘을 사용하여 요약한다. 요약을 통해 추출된 핵심 문장들은 변수에 저장되어 파이썬 API를 통해 다시 모바일에 전달된다. 마지막으로 전달된 문장들을 result 변수에 저장해 화면에 출력할 때 사용한다.

그림 4는 뉴스 기사 요약에 사용될 TextRank 알고리즘을 실행하기 위한 프로세스를 나타낸 것이다.

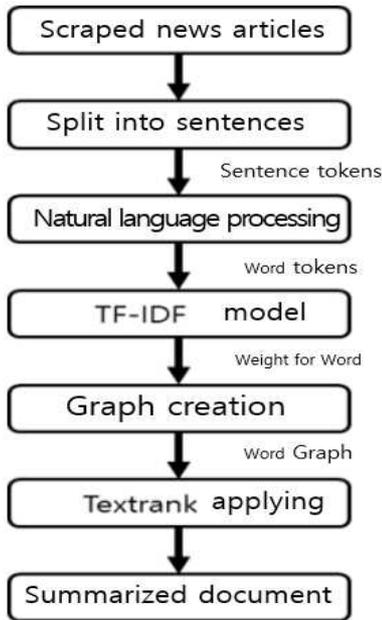


그림 4. TextRank 알고리즘 프로세스
Fig. 4. TextRank algorithm process

TextRank 알고리즘은 수집된 뉴스 기사를 먼저 문장 단위로 분리하고, 분리된 문장을 다시 자연어 처리 과정을 거치면서 명사 형태의 단어들로 추려낸다. 이 단어들을 활용해서 TF-IDF 모델을 이용해서 단어별 가중치를 매기고 그래프를 생성한다. 이 그래프에 TextRank를 적용하면 문서 내의 핵심 문장을 추출할 수 있고, 추출하고자 하는 문장의 개수는 사용자가 TextRank를 적용하기 전에 미리 설정할 수 있다. 본 논문에서는 뉴스 기사 요약 시스템에서는 핵심 문장의 개수를 3으로 설정하였다. 문장의 개수를 3으로 한 이유는 최초의 한글뉴스 요약

앱인 ‘뉴스썸머’, 빠른 정보를 얻기 위해 업계에서는 3줄요약을 많이 사용하고 있어서이다[31].

각 함수들을 이용해서 TextRank 알고리즘을 적용시키면 그래프 랭킹의 상위에 해당하는 3개의 문장을 결과로 제공하고, API를 이용해서 모바일에 전송해 사용자에게 제공한다. 표 1은 요약 기능을 구현하기 위해 사용된 TextRank 알고리즘 구현에 필요한 함수와 그 설명을 나타낸 것이다.

표 1. TextRank 알고리즘 구현에 사용되는 함수들
Table 1. Functions used to implement TextRank algorithm

Function	Description
pr()	Pagerank algorithm function
tok_sent()	Function for tokenizing sentences
vect_sent()	Function for vectorizing sentences
make_graph()	Function for generating graphs
smrz()	Function for extracting summarized sentences

Textrank 알고리즘은 문장끼리의 유사도를 담은 그래프를 가지고 Pagerank 알고리즘을 학습시켜서 문장의 랭킹을 계산하는 과정을 포함하기 때문에 pr() 함수로 Pagerank 알고리즘의 기능을 구현하였다. 이후 전체 문장에 대해서 tok_sent() 함수를 적용해서 문장을 토큰화하고, vect_sent() 함수를 통해 문장을 벡터화해서 문장 간 유사도를 계산하는데 활용하였다. 이후 make_graph() 함수를 통해 문장 간 유사도를 기반으로 문장 그래프를 생성하고 마지막으로 smrz() 함수를 통해 문장 간 유사도가 높은 문장을 핵심 문장으로 추출하였다.

먼저 데이터를 저장하기 위해서 크롤링하기 위한 뉴스 기사의 링크 주소를 가져와 변수에 저장한 다음 Jsoup.connect() 함수를 이용해서 데이터를 가져온다. 가져온 데이터들은 원하는 부분만을 파싱하여 다시 변수에 저장한다. 이렇게 저장한 뉴스 기사 데이터는 파이썬 API를 통해 파이썬 Flask 서버에 전달해서 요약을 수행한다.

애플리케이션은 파이썬으로 작성된 API와 POST 방식으로 데이터를 전송한다. 뉴스기사의 내용이 담겨있는 contents 변수를 jsonObject에 저장한 뒤 DataOutputStream을 이용해 데이터를 전송한다.

파이썬 Flask 서버에서는 request.get_json() 함수를 사용하여 모바일에서 전송한 JSON 형식의 문서 전체 내용을 담은 데이터를 전송받는다. 전송받은 데이터는 get() 함수에 변수 이름을 담아서 추출한 내용을 새로운 변수에 저장해서 요약하기 위한 파일에 전달한다.

안드로이드와 파이썬 Flask 서버 사이에 데이터를 주고받을 때는 JSON 형식의 데이터를 사용하며, 해당 데이터는 키와 값이 쌍을 이루어 사용시 키를 호출하여 값들을 변수에 저장해서 사용한다. stringBuilder 클래스를 통해 JSON 형식으로 받은 데이터를 String 형식의 데이터로 변환한 다음 화면에 출력시키기 위해 변수로 저장해서 사용했다.

실시간으로 뉴스 페이지를 모니터링하고 새로운 기사가 업로드되면 사용자에게 알리기 위해 Service를 이용하였다. Service는 사용자에게 인터페이스를 제공하지 않으며 다른 애플리케이션으로 전환해도 백그라운드에서 계속 실행된다[32].

백그라운드에서는 가장 최근에 업로드된 뉴스 기사 제목을 크롤링하여 SharedPreferences를 이용하여 변수에 저장한다. 이후 주기적으로 크롤링하여 애플리케이션에 저장된 뉴스 기사의 제목과 일치하는지를 확인하여 일치하지 않으면 알림을 통해 사용자에게 새로운 기사가 등록되었음을 알린다.

prevTitle은 마지막으로 업로드된 기사를 크롤링하여 저장하는 변수 title과 비교하기 위한 변수이다. title과 prevTitle이 일치하지 않으면 알림을 띄운 후 title 변수의 내용을 prevTitle에 저장한다.

notification을 빌드시켜 가져온 뒤 설정을 통해 액티비티가 호출 될 경우에 태스크 최상단에 존재하면 새로운 인스턴스를 생성하지 않고 존재하고 있는 액티비티를 포그라운드로 가져온다. 상단 알림창에 알림이 올 경우 상태바 드래그시 타이틀을 “새로운 기사가 등록되었습니다.”로 띄우며 서브타이틀은 설정한 title의 값을 입력한다. 알림창 터치시 알림창을 삭제하게 설정하였으며 설정한 pendingIntent로 이동하게 설정하였다.

4.2 구현화면

본 논문에서 애플리케이션은 프로토타입 형식의

로 구현하였으며, 기능을 중점적으로 구현하여 심플한 디자인으로 제작하였다. 그림 5는 뉴스 기사 목록 확인 구현 화면이다.

사용자는 먼저 업로드된 뉴스 기사의 목록을 확인할 수 있다. 또한, 화면 상단의 버튼을 터치해서 사용자가 관심 있는 분야 뉴스 기사의 새로운 소식들에 대한 알림을 받을지를 설정할 수 있다.

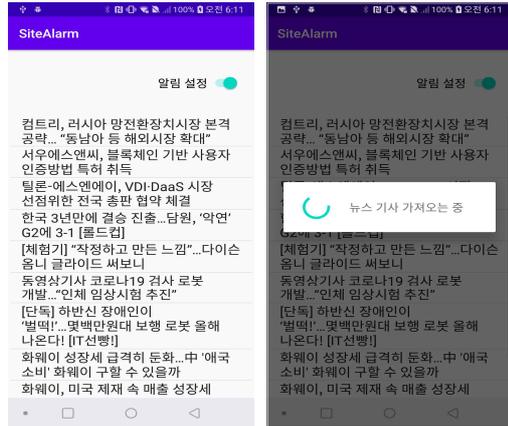


그림 5. 뉴스 기사 목록 화면
Fig. 5. News article list screen

사용자는 자신이 읽고자 하는 기사를 클릭할 수 있으며, 클릭 시 화면에 출력되는 프로그레스 바를 통해 애플리케이션에서 해당 뉴스 기사를 불러오고 있다는 것을 인지할 수 있고 해당 뉴스를 볼 수 있는 페이지로 이동한다.

그림 6은 사용자가 읽고자 하는 문서를 클릭한 경우 볼 수 있는 화면이다. 이 화면에서 사용자는 문서의 요약 기능을 실행할 수 있으며, 요약된 문서 내용을 제공받을 수 있다.

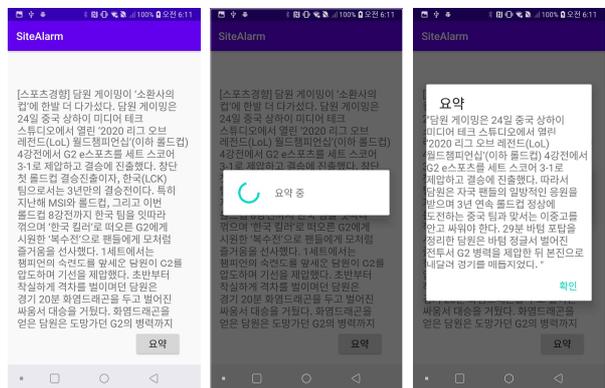


그림 6. 문서 요약 수행 과정 및 결과 화면
Fig. 6. Document summary process and results screen

시스템에서는 문서의 본문에 요약 버튼을 삽입해서 사용자가 원할 경우에만 문서의 요약 내용을 확인할 수 있도록 하였다. 그리고 문서 요약이 수행될 때는 프로그레스 바를 통해 사용자에게 시스템이 제대로 작동하고 있다는 것을 인지시키는 역할을 할 수 있도록 추가하였고, 문서 요약이 모두 완료됐을 때엔 팝업 창에 요약된 내용을 첨부해서 사용자가 간단하게 문서의 요약 내용을 확인할 수 있다.

그림 7은 사용자들에게 더욱 편리한 서비스를 제공하기 위해서 새로운 뉴스 기사가 등록되었을 때 알림 설정을 한 사용자에게 한해서 뉴스 기사에 대한 알림이 가도록 설정한 화면이다.

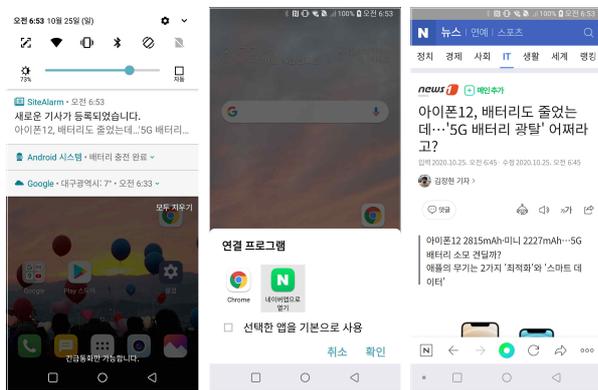


그림 7. 새로운 뉴스 기사 알림 화면
Fig. 7. New news article notification screen

사용자는 자신이 원하는 기사가 등록되면 알림 설정을 통해 자동으로 알림을 받아 모바일 상단 바에서 기사의 제목을 확인할 수 있고, 내용을 확인하고자 하는 기사에 대해서는 기사 알림을 터치했을 경우에 본문으로 바로 이동해서 내용을 확인하고, 바쁜 경우에는 요약 버튼을 활용해서 빠르게 문서의 전체 내용을 요약하여 확인할 수 있다. 알림의 내용은 애플리케이션에서 URL 기사내용을 크롤링한 후 ‘알림’ 버튼을 클릭하면 파이썬 API로 JSON 데이터를 전송하고, 전송받은 JSON 데이터를 파싱하여 TextView에 표시하여 애플리케이션 알림기능으로 나타낸다.

본 논문의 추출요약 기법을 사용한 뉴스 기사 요약 어플리케이션은 기존 논문들에서 정적인 요약 결과만을 제공한 것과 달리 사용자가 원하는 경우에만 요약을 확인 가능하도록 선택적 요약 기능을

추가하고, 새로운 기사 업로드시 사용자에게 알림을 제공하여 사용자 맞춤형 애플리케이션으로 설계하고 구현하였다.

V. 결 론

본 논문에서는 뉴스 기사를 요약하여 정보를 빠르게 확인하기 위하여 파이썬에서 제공하는 JSON 데이터 전송을 지원하는 API 연동을 사용하여 서버와 모바일 간 통신을 하였다. 그리고 서버에서 뉴스 기사를 요약하고 요약된 뉴스 기사 내용을 다시 모바일에 전송하여 사용자에게 제공하는 애플리케이션을 구현하였다. 또한 사용자들이 빠르게 기사 내용을 파악할 수 있도록 Textrank 알고리즘을 적용하여 내용을 요약하였다. 알림 설정 기능을 추가해서 새로운 뉴스기사가 등록되었을 때 사용자에게 뉴스 기사의 제목을 제공해서 기사를 더욱 신속하게 전달할 수 있도록 하였다.

본 논문의 연구는 사람이 내용을 요약하는 것처럼 완전히 새로운 문장으로 요약하는 것이 아닌 문장 자체를 추출하는 알고리즘을 사용하여 추출에 가깝다고 할 수 있다. 향후 연구로는 실제로 문장을 생성해서 요약하는 생성 요약 방법과 본 논문의 시스템에 사용된 추출 요약을 병행해서 두 가지 방법 중 더 만족도가 높은 요약 방법을 채택하는 사용자에게 요약된 내용을 제공하는 시스템을 개발하기를 기대한다.

References

- [1] Y. C. Jung and Y. H. Kim, "2020 Broadcast Media Usage Behavior Survey", Korea Communications Commission, pp. 83-87, Dec. 2020.
- [2] Korea Press Foundation, "Media Users in Korea 2020", Korea Press Foundation, pp. 63, Dec. 2020.
- [3] D. W. Kim, "Big Data Use Cases of the Sector", Dong-A University Business Research Center, The Journal of Management, Vol. 34, pp. 39-52, Dec. 2013.

- [4] J. W. Jeon, H. S. Hwang, and C. K. Lee, "Two-step Document Summarization using Extractive Summarization and Abstractive Summarization", *The Korean Institute of Information Scientists and Engineers*, Vol. 46, No. 1, pp. 581-583, Jun. 2019.
- [5] H. C. Kim and S. W. Lee, "Korean Text Automatic Summarization using Semantically Expanded Sentence Similarity", *Proceedings of the Korea Information Processing Society Conference*, Vol. 21, No. 2, pp. 841-844, Nov. 2014. <https://doi.org/10.3745/PKIPS.y2014m11a.841>.
- [6] H. C. Kim and S. W. Lee, "Graph-based Automatic Extractive Summarization on Multi-document", *Proceedings of the Korean Society of Information Sciences Academic Presentation*, pp. 778-780, Dec. 2015.
- [7] C. S. Ko, "The Prospects of Natural Language Process", *Urimal*, No. 31, pp. 5-22, Oct. 2012.
- [8] H. Saggion and T. Poibeau, "Automatic text summarization: Past, present and future", In *Multi-source, multilingual information extraction and summarization*, pp. 3-21, 2013.
- [9] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization", *Association for Computational Linguistics, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379-389, Sep. 2015. <https://doi.org/10.48550/arXiv.1509.006853>.
- [10] W. H. Lee, S. I. Cho, and D. H. Kim, "Proposal of keyword extraction method based on morphological analysis and PageRank in Tweeter", *Journal of Digital Contents Society*, Vol. 19, No. 1, pp. 157-163, Jan. 2018. <http://dx.doi.org/10.9728/dcs.2018.19.1.157>.
- [11] C. Mallick, A. K. Das, M. Dutta, A. K. Das, and A. Sarkar, "Graph-based text summarization using modified TextRank", *Soft Computing In Data Analytics*, pp. 137-146, Aug. 2019. https://doi.org/10.1007/978-981-13-0514-6_14.
- [12] J. K. Lee, S. B. Park, and S. J. Lee, "Meeting Minutes Summarization using Two-step Sentence Extraction", *Journal of Korean Institute of Intelligent Systems*, Vol. 20, No. 6, pp. 741-747, Dec. 2010. <https://doi.org/10.5391/JKIIS.2010.20.6.741>.
- [13] G. H. Kim and Y. J. Ko, "Query-based Abstractive Summarization Model Using Sentence Ranking Scores and Graph Techniques", *Journal of KIISE*, Vol. 47, No. 12, pp. 1172-1180, Dec. 2020. <http://dx.doi.org/10.5626/JOK.2020.47.12.1172>.
- [14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks", *Advances in neural information processing systems*, pp. 3104-3112, Dec. 2014.
- [15] K. H. Choi and C. K. Lee, "End-to-end Korean Document Summarization using Copy Mechanism and Input-feeding", *Journal of KIISE*, Vol. 44, No. 5, pp. 503-509, May 2017. <https://doi.org/10.5626/JOK.2017.44.5.503>.
- [16] T. Y. Kim, D. W. Kwon, H. W. Kim, and H. T. Ju, "State of Art on Open API Development", *KNOM Review*, Vol. 18, No. 1, pp. 25-34, Aug. 2015.
- [17] H. S. An, B. K. Park, Y. C. Kim, and J. H. Lee, "Research practices on Integrated Architecture based on REST API", *Journal of Platform Technology*, Vol. 7, No. 1, pp. 3-9, Apr. 2019.
- [18] TigerCow.Door, <https://doorbw.tistory.com/31> [accessed: May 15, 2024]
- [19] J. H. Ko, Y. C. Kim, and H. W. Lee, "Smart Home Automation using Arduino and Python Web Server", *The Korean Institute of Information Scientists and Engineers*, pp. 1820-1822, Jun. 2016.
- [20] D. M. Seo and H. M. Jung, "Intelligent Web Crawler for Supporting Big Data Analysis Services", *The Journal of the Korea Contents Association*, Vol. 13, No. 12, pp. 575-584, Dec. 2013. <https://doi.org/10.5392/JKCA.2013.13.12.575>.

[21] S. H. Park, Y. H. Choi, J. O. Lee, and J. S. Seo, "Automatic Summarization of Text by Semantic Filtering: Context Search System", Korean French Studies, Vol. 47, Aug. 2004.

[22] E. K. Cho, "Development of Sentence Compression Method for Automatic Text Summarization: Using Sejong syntactic analysis corpus", The Korean Language And Culture Education Society, Vol. 14, No. 3, pp. 201-220, Aug. 2018. <http://dx.doi.org/10.18842/klaces.2018.14.3.9>.

[23] H. G. Kim, "Efficient Keyword Extraction from Social Big Data Based on Cohesion Scoring", Journal of the Korea Society of Computer and Information, Vol. 25, No. 10, pp. 87-94, Oct. 2020. <http://dx.doi.org/10.9708/jksci.2020.25.10.087>.

[24] J. K. Hong, J. Lee, and W. S. Lee, "An Associated Keywords Extraction and a Spread Clustering Methods for an Efficient Document Search", Journal of KIIT, Vol. 9, No. 6, pp. 155-166, Sep. 2011.

[25] S. H. Kim, S. Yang, and Y. J. Ko, "Extracting Korean-English Parallel Sentences from Wikipedia", Journal of KISS : Software and Applications, Vol. 41, No. 8, pp. 580-585, Aug. 2014.

[26] D.-H. Kim, S.-Y. Yun, S.-Y. Chae, and S.-J. Yu, "A News Recommendation System using Text Mining and LOD Techniques", Journal of KIIT, Vol. 20, No. 8, pp. 1-6, Aug. 2022. <http://dx.doi.org/10.14801/jkiit.2022.20.8.1>.

[27] Y. J. Bae, H. T. Jang, T. W. Hong, and H. Y. Lee, "Automatic Meeting Summary System using Enhanced TextRank Algorithm", Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol. 11, No. 5, pp. 467-474, Oct. 2018. <https://doi.org/10.17661/JKIIECT.2018.11.5.467>.

[28] Y. S. Cho, K. H. Ahan, and S. K. Kim, "A Study on Key Sentence Extraction by using Enriched TextRank", Korean Institute of

Information Technology, Proceedings of KIIT Conference, pp. 39-44, May 2013.

[29] Y. D. Kwon, N. R. Kim, and J. H. Lee, "Document Summarization Considering Entailment Relation between Sentences", Korean Institute of Information Scientists and Engineers, Journal of KIISE, Vol. 44, No. 2, pp. 179-185, Feb. 2017. <https://doi.org/10.5626/JOK.2017.44.2.179>.

[30] Android Developers, <https://developer.android.com/guide/topics/ui/notifiers/notifications>. [accessed: Oct. 12, 2021]

[31] <https://www.mediatoday.co.kr/news/articleView.html?idxno=110676>. [accessed: May 15, 2024]

[32] Android Developers, <https://developer.android.com/guide/components/services> [accessed: Oct. 12, 2021]

저자소개

문 봉 관 (Bongkwan Moon)



2020년 2월 : 국립안동대학교
멀티미디어공학과(공학사)
2022년 2월 : 국립안동대학교
멀티미디어공학과(석사)
관심분야 : 인공지능, 블록체인,
빅데이터 분석

임 한 규 (Hankyu Lim)



1981년 : 경북대학교 전자공학과
(공학사)
1984년 : 연세대학교
컴퓨터공학과(공학석사)
1997년 : 성균관대학교
컴퓨터공학과(공학박사)
1981년 ~ 1982년 : 대한주택공사

연구원

1982년 ~ 1986년 : ETRI 위성통신연구실 연구원
1986년 1994년 : 한국IBM 소프트웨어연구소 연구원
1994년 ~ 1998년 : 한서대학교 전산정보학과 교수
1998년 ~ 현재 : 안동대학교 소프트웨어융합학과 교수
관심분야 : 멀티미디어 콘텐츠, 웹 애플리케이션, NLP

박 은 주 (Eunju Park)



1993년 : 안동대학교 전산통계학과
(이학사)

2001년 : 안동대학교 대학원
컴퓨터공학과(공학석사)

2016년 : 안동대학교 대학원
정보통신공학과 멀티미디어 공학
전공(공학박사)

2000년 ~ 2016년 : 안동대학교 교양교육원 강사

2016년 ~ 2017년 : 안동대학교 멀티미디어공학과
강의전담교원

2019년 ~ 현재 : 안동대학교 SW융합교육원 교수

관심분야 : 웹 접근성, 모바일 접근성, 유니버설 디자인