

스마트 전자게시판을 위한 학생 맞춤형 게시물 표출 시스템 개발

김태형*, 김민영**¹, 김주엽**², 전지용**³, 정은혜**⁴, 현소담**⁵

Development of a Student-customized Post Display System for Smart Electronic Bulletin Boards

Tae-Hyong Kim*, Min-Young Kim**¹, Ju-Yeop Kim**², Ji-Yong Jeon**³, Eun-Hye Jeong**⁴, and So-Dam Hyeon**⁵

이 연구는 2021년 국립대학 육성사업비로 지원되었음

요약

본 연구는 학내외 게시물 및 뉴스를 학생 맞춤형으로 선별하고 이를 웹 기반으로 제공하여 스마트 전자게시판에 활용할 수 있는 시스템을 제안한다. 학생들은 이 시스템을 통해 키오스크에서 RFID 또는 QR 코드 리더에 학생증을 인식하여 자신에 맞추어 선별된 게시물이나 뉴스를 확인할 수 있다. 제안 시스템은 학생을 조건에 따라 특정 범주로 나누어 각 범주에 해당하는 키워드를 사전에 추출하고, 학내외 최신 게시물을 매일 크롤링한 후 키워드를 추출하여 게시물과 함께 저장한다. 시스템에 학번이 입력되면 해당 학생의 범주 키워드와 각 게시물의 키워드 사이의 연관성을 측정하여 유사도가 높은 게시물을 추천 게시물로 정한다. 또한 추천 게시물을 전자게시판에 적절히 표출하기 위해 텍스트 요약 및 포맷팅을 수행한다. 개발된 시스템을 실제 학내 전자게시판에 설치하여 시험 운용한 결과 유용성을 확인할 수 있었다.

Abstract

This study proposes a system that can be used for smart electronic bulletin boards by selecting posts and news on campus and off campus and providing them on the web. Through this system, students can check the posts or news selected for them by scanning their student ID card with an RFID or QR code reader at a kiosk. The proposed system divides students into specific categories based on their conditions, extracts keywords corresponding to each category in advance, crawls the latest on-campus and off-campus posts daily, extracts the keywords, and stores them with the posts. When a student's ID is given into the system, it measures the relevance between the student's category keywords and the keywords in each post and recommends posts with high similarity. It also performs text summarization and formatting to properly display the recommended posts on the bulletin board. The developed system was piloted on a real campus electronic bulletin board and its usefulness was confirmed.

Keywords

smart electronic bulletin boards, post recommendation, keyword matching, text summarization, smart campus

* 금오공과대학교 인공지능공학과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0003-3806-2517>

** 금오공과대학교 컴퓨터공학과

- ORCID¹: <https://orcid.org/0009-0009-4769-7937>

- ORCID²: <https://orcid.org/0009-0008-9743-6949>

- ORCID³: <https://orcid.org/0009-0003-3850-3180>

- ORCID⁴: <https://orcid.org/0009-0009-9529-7903>

- ORCID⁵: <https://orcid.org/0009-0006-9269-1540>

• Received: Feb. 02, 2024, Revised: Feb. 23, 2024, Accepted: Feb. 26, 2024

• Corresponding Author: Tae-Hyong Kim

Dept. of Artificial Intelligence Engineering, Kumoh National Institute of Technology, 61 Daehak-ro, Gumi, Gyeongbuk, Korea

Tel.: +82-54-478-7528, Email: taehyong@kumoh.ac.kr

I. 서 론

최근 4차산업혁명 기술의 비약적 발전으로 인공지능, 빅데이터, 클라우드, 사물인터넷 관련 국내외 시장이 빠르게 성장하고 있으며 이러한 기술에 기반한 다양한 지능형 서비스가 출시되고 있다. 이러한 변화는 교육환경에도 나타나고 있는데 COVID-19로 인한 비대면 교육의 확산을 계기로 5G 초고속 무선 통신 및 메타버스 등을 이용한 스마트 캠퍼스 구축으로 발전하고 있다[1]. 교육 및 학습 환경 고도화 등의 인프라 구축과 더불어 인공지능 및 빅데이터를 이용한 학생 맞춤형 서비스의 제공은 스마트 캠퍼스 구축의 중요한 요소이다. 최근 대학들은 인공지능 기반으로 자신의 핵심 역량을 진단하고, 수강 이력과 진로 목표에 적합한 교과목 및 비교과 활동을 추천하는 맞춤형 교육정보 서비스를 도입하고 있다[2]. 이러한 학생 맞춤형 서비스는 아직 주로 교육 분야에 한정되어 있어 학생들에게 도움을 줄 수 있는 다양한 분야의 서비스 발굴이 요구되고 있다.

학생들이 대학 생활에서 어려움을 겪는 것 중 하나는 매일 같이 쏟아지는 수많은 정보 중에서 자신에게 필요한 것을 파악하는 것이다. 학내에서는 각 부서와 학과에서 수많은 공지를 다수의 개별 홈페이지에 올리고 있으며 다양한 외부 매체에서도 전공과 관련된 정보와 뉴스가 끊임없이 생산되고 있어서 이러한 정보들을 모두 파악하여 자신에게 필요한 것을 제때 숙지하기는 매우 어려운 일이다. 캠퍼스 내에 대형 디스플레이를 설치하여 실시간으로 학생들에게 주요 정보를 제공하기도 하지만 맞춤형 정보가 아니다 보니 활용도와 효용성은 그다지 높지 않은 실정이다. 따라서 학생 맞춤형으로 학내외 정보를 제공할 수 있는 지능형 데이터 큐레이션 서비스 개발이 요구되고 있다.

본 연구는 대학 내에 학내외 게시물 및 뉴스 정보를 학생 맞춤형으로 선별하고 이를 웹 기반으로 제공하여 스마트 전자게시판에 활용할 수 있는 시스템을 개발하였다. 학생들은 키오스크에서 자신의 학번을 입력하거나 RFID(Radio-Frequency Identification) 또는 QR(Quick Response) 코드 리더에 학생증을 인식하여 자신에게 맞춤형으로 선별된 게시물이나 뉴스를 확인할 수 있다.

본 논문의 구성은 다음과 같다. II장에서 기존의 관련 연구를 살펴본 후 III장에서 제안 시스템의 설계 방향과 전체 구조를 소개하고 IV장에서 시스템의 서비스 구현을 위한 각 모듈의 개발 방법을 자세히 설명한다. V장에서는 구현 방법 및 결과를 제시하고 마지막 VI장에서 결론 및 향후 과제에 관하여 기술한다.

II. 관련 연구

최근 발표된 웹 크롤링 기반의 맞춤형 뉴스 추천 연구는 주로 개인 맞춤형 뉴스 추천 시스템에 관한 것이다[3]-[6]. 이 연구들은 개인에게 적합한 뉴스를 선별하기 위해 웹 브라우저의 쿠키 정보 등을 통해 수집한 기사 조회 기록을 사용한다. 또한 해당 기사의 키워드를 추출하기 위해 TF-IDF(Term Frequency-Inverse Document Frequency) 가중치를 사용하였다. 매칭 적합도가 높은 기사를 선별하기 위해서는 키워드 매칭[3][5], 벡터공간 유사도[4], SVD(Singular Value Decomposition) 기반 유사도[6] 등이 사용되었다. 추가로 뉴스에 가중치를 부여하기 위해 조회수, 댓글수, 좋아요 수, 게시 시각 등을 활용하였다.

D.-H. Kim et al.[7]은 크롤링된 기사에 대해 그래프 기반으로 중요도를 계산하는 TextRank 모델[8]을 사용하여 일반적인 뉴스 추천 시스템을 개발하되, 군집화(Clustering)를 이용한 계층적 추천 방식을 제시하였다. K. Sin et al.[9]은 특정 분야의 사용자에게 게시물을 추천하기 위해 사용자가 사전에 등록한 키워드를 활용하는 토픽 모델링을 사용하였다.

제안하는 시스템은 여러 사람이 같이 볼수 있는 스마트 전자게시판에 추천 게시물 및 뉴스를 표출하기 때문에 뉴스 조회 기록과 같은 개인 정보 기반의 추천 방식 대신, 학생을 범주화하여 각 범주별 키워드를 사전에 추출하는 토픽모델링 기법을 사용한다.

III. 제안 시스템

3.1 시스템 설계

제안하는 시스템에서 학생 맞춤형으로 선별하고자 하는 정보는 대학과 학과 홈페이지에 존재하는

다양한 게시물, 국내 일반 언론사와 전공 관련 전문 언론사의 뉴스 기사 및 취업 관련 사이트의 게시물을 대상으로 하였다. 향후 전공 또는 청년과 관련된 다양한 외부 기관 홈페이지에 게시되는 정보까지 그 범위를 확장할 수 있을 것이다.

학생들은 맞춤형 서비스를 제공받기 위해 자신의 학번을 입력하여야 한다. 사용 편의성을 위해 서비스는 웹 기반으로 사용할 수 있도록 설계하였으나 시험 서비스 단계에서는 서비스 부하를 고려하지 않은 가벼운 웹서버 모듈을 활용하기 때문에 웹주소를 공개하지 않고 대형 전자게시판에 학생증의 ID 정보를 인식시켜 서비스를 이용할 수 있도록 RFID 및 QR 코드 리더를 갖춘 키오스크(Kiosk)를 설치하였다. 학생들이 자신의 ID를 이용해 로그인 하면 대형 전자게시판을 통해 자신에게 맞춤 선별된 학내외 게시물을 확인할 수 있다. 또한 표시되는 게시물 글자가 크고 화면이 계속 전환되는 전자게시판의 특성상 현장에서 모든 정보를 확인하기는 어렵기 때문에 키오스크를 통해 게시물의 상세 정보 URL(Uniform Resource Locator)을 휴대전화로 전송받을 수 있도록 하였다.

학생 맞춤형 게시물을 실시간으로 표출하기 위해서는 학번 정보를 입력받아 그때그때 검색을 통해

결과를 보여주기는 어렵다. 따라서 학내외 최신 게시물을 매일 크롤링한 후 키워드를 추출하여 게시물과 함께 데이터베이스에 저장한다. 또한 학과, 학년, 성별, 학적 상태, 병적 사항 등의 학생 정보를 범주화하고 각 범주에 해당하는 키워드를 사전에 추출한다. 학번이 입력되면 이 정보를 바탕으로 해당 학생의 범주에 해당하는 키워드와 사전에 크롤링(Crawling)된 기사의 키워드 사이의 연관성을 측정하고 유사도가 높은 순위로 정렬하여 선별할 게시물을 결정한다.

그런데 게시물의 다양한 특성으로 인해 선별된 게시물을 그대로 전자게시판에 표출하기는 어렵다. 전자게시판은 비교적 떨어진 거리에서 화면을 보게 되고 한 화면에 하나의 게시물만 보여주는 것이 일반적이므로 시인성을 높이기 위해 게시물을 정해진 형식에 맞게 가공해야 한다. 개발하는 학생 맞춤형 게시물 표출 시스템의 서비스 흐름도는 그림 1과 같다.

3.2 시스템 구성

학생 맞춤형 게시물 표출 시스템의 구조는 그림 2와 같다.

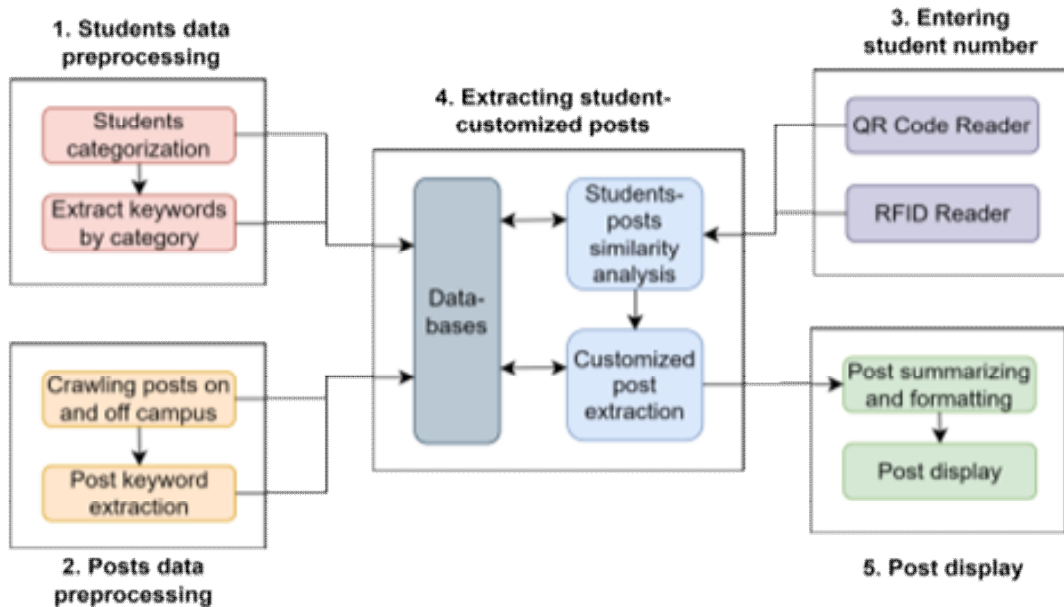


그림 1. 학생 맞춤형 게시물 표출 시스템 흐름도
Fig. 1. Personalized post display system flowchart

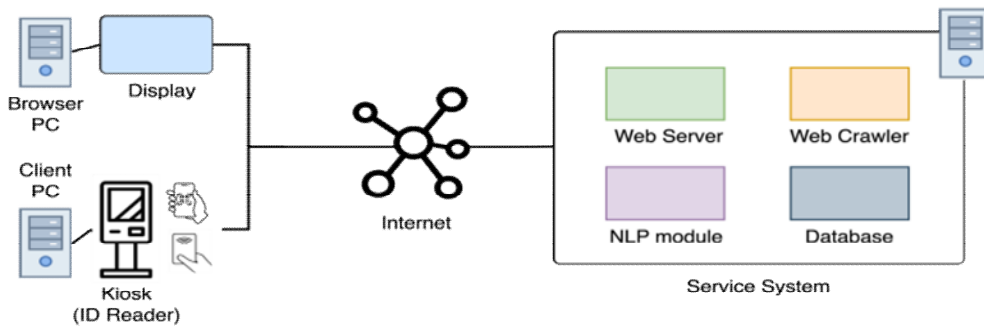


그림 2. 학생 맞춤형 게시물 표출 시스템의 구조
Fig. 2. Structure of the student-customized post display system

서비스 시스템은 웹 기반 서비스를 위한 웹서버 모듈과 학내외 게시물을 크롤링하기 위한 웹 크롤러 모듈, 텍스트 데이터 처리를 위한 자연어처리(NLP, Natural Language Processing) 모듈 및 데이터베이스로 구성된다. 웹서버는 시험용 시스템에서는 파이썬 기반의 경량 웹 프레임워크인 Flask를 사용하고, 웹 크롤러는 정적 크롤링을 위한 파이썬 기반의 BeautifulSoup 라이브러리와 동적 크롤링을 위한 웹 어플리케이션 자동화 및 테스트 프레임워크인 Selenium 라이브러리를 사용한다. NLP 모듈은 역시 파이썬 기반의 한글 자연어처리 라이브러리인 KoNLPy와 토픽 모델링 라이브러리 Gensim을 사용한다. 또한 데이터베이스는 다양한 형태의 데이터 저장이 가능한 NoSQL 기반의 문서 지향적 데이터베이스인 MongoDB를 사용한다. 웹 프론트엔드 개발에는 속도가 빠르고 경량인 Vue.js 프레임워크와 Axios HTTP 통신 라이브러리를 사용한다.

사용자가 접근하는 키오스크에는 RFID 및 QR 코드 리더로부터 획득한 ID 정보를 전송하고, 서비스 시스템에서 추출된 게시물을 수신하기 위해 키오스크 내 클라이언트 PC에서 Requests HTTP 라이브러리를 사용한다. 포맷팅된 게시물은 전자게시판에 연결된 웹 브라우저 PC에서 정해진 URL 주소로 접속하여 표출된다.

IV. 개발 내용

4.1 RFID/QR 코드 입력 인터페이스

키오스크에는 RFID 및 QR 코드 리더와 사용자 조작 용 소형 디스플레이가 클라이언트 PC에 연결

된다. 클라이언트 소프트웨어는 GUI(Graphic User Interface) 화면을 통해 사용자가 직접 학번을 입력하거나 RFID 및 QR 코드를 인식시킬 때까지 대기한다. RFID 리더는 학생증에 내장된 RFID 칩을 인식하여 개인 고유번호를 추출하고, QR 코드 리더는 휴대전화에 설치된 전자학생증 QR 코드를 읽어 학번 정보를 추출한다. 클라이언트 소프트웨어가 RFID 및 QR 코드 리더 API(Application Programming Interface)를 이용해 인식된 정보를 추출하면 REST (Representation State Transfer) API를 이용해 서비스 시스템에 전송한다. 사용하는 RFID 리더는 USB 인터페이스로 PC에 연결되며 13.56MHz 무선 통신을 지원하고, QR 코드 리더는 USB-RS232 통신 방식으로 PC에 연결되며 초당 100 장의 648×488 해상도 스캔을 지원한다. 그림 3은 키오스크에 설치된 QR 코드 리더(좌측)와 RFID 리더(우측)를 보여준다.

키오스크 화면은 대기상태에서 입력 칸과 로그인 버튼을 표시하고 로그인이 된 후에는 로그인한 사용자 정보를 출력한다. 이후 일정 시간이 지나면 로그인된 사용자는 자동 로그아웃되고 화면은 대기상태로 전환된다.



그림 3. 설치된 RFID (좌측) 및 QR코드 리더 (우측)
Fig. 3. Installed RFID (left) and QR code reader (right)

4.2 교내외 정보 수집

교내외의 게시물 정보는 웹 크롤링을 사용하지 않고 대학 정보전산원의 협조를 받아 각 부서와 학과 홈페이지의 게시글 데이터베이스에 직접 접근하여 수집할 수 있었다. 교내 수집 게시글은 대학 메인 홈페이지의 학사 안내, 행사 안내 및 일반 소식과 학과별 홈페이지의 공지 사항, 취업 정보를 우선 수집 대상으로 하였다. 교내 게시글의 수집 정보는 최근 1주일간의 게시물에 대한 제목, 작성자, 등록일, 조회수 및 본문 내용이다.

교외 게시물 정보는 크롤링 접근이 쉬운 뉴스 기사와 취업 및 공모전 정보를 대상으로 하였다. 뉴스 기사는 KBS 뉴스[10], 전자신문[11], AI타임스[12]를 우선 선정하였고 향후 수집 대상을 확대할 예정이다. 취업 정보와 공모전 정보는 각각 잡코리아[13]와 썬긱[14]을 수집 대상으로 선정하였다. 크롤링 데이터 범위는 게시글 제목과 부제목, 날짜, 등록일, 공모전 진행 기간, 채용 모집 마감일, 웹주소, 본문 및 이미지 소스 등이다. 뉴스 기사는 언론사의 뉴스에 대한 통합 조회 기능을 제공하고 있는 네이버 뉴스[15]를 활용하여, 최근 1주간의 기사 목록 정보를 통해 언론사의 뉴스 기사 페이지로 접근하고 기사 내용을 크롤링한다. 그림 4는 KBS 뉴스의 개별 기사 페이지에서 크롤링하는 정보를 보여준다.



그림 4. 뉴스 기사에서의 크롤링 정보[10]
Fig. 4. Crawling information from a news article[10]



그림 5. 채용 사이트에서의 크롤링 정보[13]
Fig. 5. Crawling information from a recruitment site[13]

썬긱의 경우 개별 게시글 접근 없이 전체 공모전 게시판 테이블에서 공모 명, 웹주소, 카테고리, 주최, 진행 기간 등의 데이터를 수집하고 카테고리 및 주최기관 등의 내용을 텍스트로 통합하여 본문 내용으로 전송한다. 잡코리아 역시 개별 게시글 접근 없이 전체 게시글 검색 페이지에서 게시글 제목, 웹주소, 모집 기간, 카테고리, 회사정보, 지원 자격, 고용 형태, 태그 정보 등의 데이터를 수집하고 회사정보와 지원 자격, 고용 형태 및 태그 정보를 텍스트로 통합하여 본문 내용으로 전송한다.

수집하는 정보의 최신성을 확보하기 위해 매일 크롤링이 실행되며, 사용자의 이용이 현저히 적은 심야 시간을 이용하여 자동으로 실행되도록 스케줄링하였다.

4.3 학생 범주별 맞춤 추천 키워드 추출

학생에게 적합한 게시물을 신속히 검색하기 위해서 학생을 범주화하고 그 범주에 대한 검색 키워드를 추출한다. 학생 범주화에 사용된 정보는 소속 학과, 학년, 성별, 군필 여부, 학적 상태 등이다. 해당 정보는 교내 정보전산원의 협조를 받아 학사 정보 데이터베이스로부터 수집하였다.

각 학생 범주에 맞는 적절한 검색 키워드를 추출하는 것은 추천 게시물 매칭 정확도에 큰 영향을 끼치므로 매우 중요한 과정이다. 따라서 검색 키워드를 객관적으로 추출하기 위해서 빅카인즈의 뉴스 검색 및 분석 시스템[16], 네이버 키워드 트렌드[17], 학내 실시간 인기 검색어[18], 한국지능정보사회진흥원의 SNS 기반 소셜 데이터 분석 시스템[19]

등을 활용하였다. 이중 뉴스와 키워드의 수집은 빅 카인즈를 주로 활용하였고 네이버 등 나머지 다른 사이트는 수집된 키워드의 유효성을 확인하기 위해 보조적으로 사용하였다. 뉴스는 검색 날짜 기준 1년 내의 모든 뉴스 중 최신순으로 최대 2만건을 대상으로 하고 뉴스의 기본정보 및 계층별 범주 정보 등의 메타데이터와 인물, 위치, 기관 및 키워드 정보 등을 분석에 사용하였다.

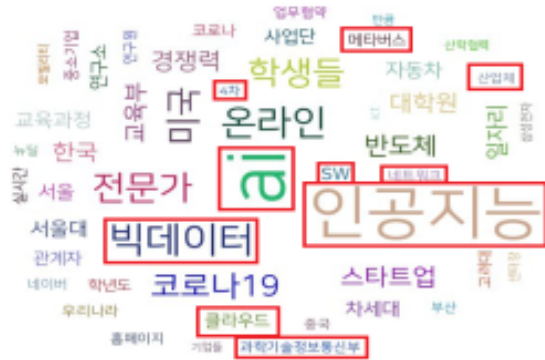


그림 6. 컴퓨터공학과 범주의 워드 클라우드 예시
Fig. 6. Word cloud example of computer engineering department category

표 1. 각 학년 범주의 대표 키워드 예시

Table 1. Examples of representative keywords for each grade category

Category	Representative keywords
1st grade	new student, first grade, scholarship, tuition, university, COVID-19, online, current student, certificate, admission
2nd grade	2nd grade, scholarship, university, COVID-19, online, current student, certificate, career path, mentor, program
3rd grade	3rd grade, scholarship, university, COVID-19, online, current student, certificate, career path, mentor, program
4th grade	4th grade, scholarship, COVID-19, online, current student, employment, graduate school, job, program, recruitment

뉴스 기사부터 각 학생 범주별 대표 키워드를 선택하기 위해 워드 클라우드 기법을 사용하였다. 각 범주의 이름 및 유사어를 검색식으로 구성하여 뉴스 기사를 검색한 결과에서 키워드 가중치와 빈도수를 계산하여 워드 클라우드를 생성한다. 생성된 워드 클라우드에서 의미 있는 키워드를 추출하고, 다시 추출된 키워드와 기존 검색식을 조합하여 새로운 검색식을 도출하는 과정을 반복하여 각 학생 범주의

대표 키워드를 결정한다. 대표 키워드는 맞춤 게시물의 수를 적절히 조절하기 위해 최소 5개에서 최대 10개까지로 제한을 두었다. 컴퓨터공학과 범주의 대표 키워드로는 '인공지능', '4차산업', '메타버스', '소프트웨어', '개발', '클라우드', '데이터베이스', '블록체인', '운영체제'가 선정되었다. 그림 6과 표 1은 각각 '컴퓨터공학과' 범주의 워드 클라우드 예시와 각 학년 범주에 해당하는 대표 키워드를 보여준다.

4.4 교내외 정보 맞춤 추천

게시물의 빠른 검색 및 유사도 분석을 위해서는 게시물 문서의 키워드를 사전에 추출해야 한다. 먼저 게시물 텍스트의 형태소 분석을 통해 명사를 추출한 다음 RAKE(Rapid Automatic Keyword Extraction)[20]를 이용해 명사 텍스트로부터 키워드를 선별한다. RAKE 알고리즘은 문자열의 개수, 어구에 포함되는 단어들의 최대 개수, 전체 문서에서 단어가 몇 번 등장하는가를 기준으로 키워드 후보군을 추출하고, 후보 키워드에 나타나는 단어의 동시 발생을 사용하여 점수를 부여한다. 즉, 동시 출현한 횟수가 많은 단어가 중요한 의미가 있는 키워드가 된다. 마지막으로 추출된 키워드에서 불용어(Stopword)를 제거한다. 불용어는 일반적으로 사용하는 678개의 한글 불용어 목록을 사용했다. 그림 7은 각 학생 범주(좌)와 게시물 문서(우)로부터 키워드를 추출하는 과정을 요약해 보여준다.

학생 범주 키워드와 게시물 문서 키워드 사이의 유사도 분석을 위해서는 일반적으로 두 키워드를 같은 벡터 공간에 위치시키고 두 벡터 간 유사한 정도를 측정하게 된다. 단어를 벡터로 변환하는 워드 임베딩(Word embedding) 기법 중 흔히 사용하는 정적 기법으로는 Word2Vec[21], Fasttext[22], Glove[23] 등이 있다. 정적 워드 임베딩은 단어와 벡터값이 1:1로 고정되어 있어서 문맥에서의 의미가 고려되지 못하는 단점이 있는 반면에 최신 동적 워드 임베딩은 문맥에 따라 실시간으로 임베딩 벡터를 생성해야 하므로 실행시간이 더 많이 소요된다. 본 시스템에서는 많은 게시물에 대한 유사도를 빠르게 측정해야 하므로 대표적인 정적 워드 임베딩 기법인 Word2Vec을 사용했다.

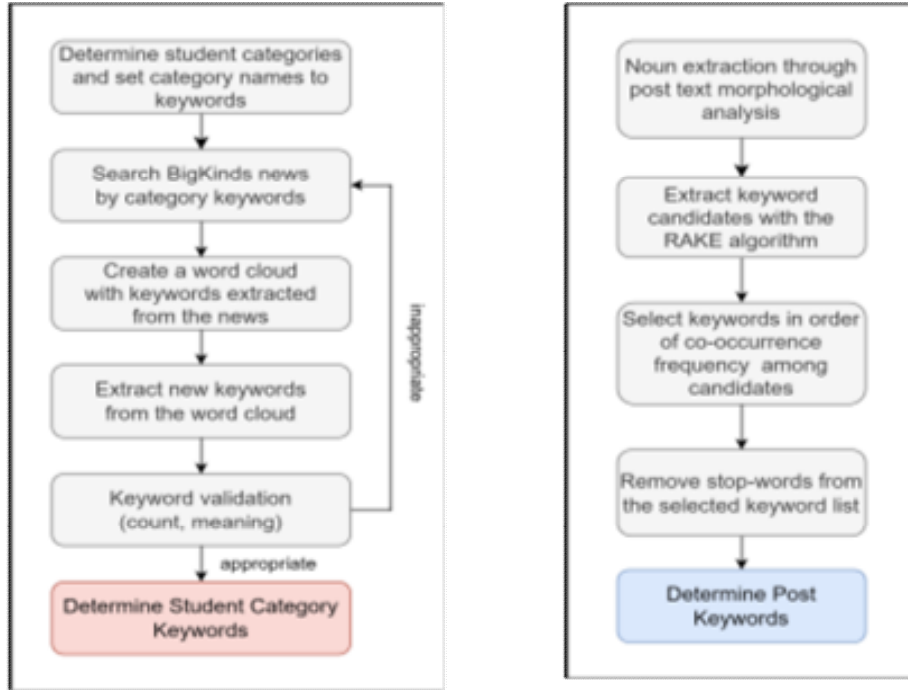


그림 7. 학생 범주와 게시물 문서에서 키워드 추출 과정
 Fig. 7. Keyword extraction process from student categories and post documents

사용한 워드 임베딩 모델은 Gensim 라이브러리에서 제공하는 한국어 사전학습 임베딩 모델로 벡터 크기 200차원, 학습 단어 수 30,185개, 학습 코퍼스 크기 339MB인 모델이다. 사용자 u 와 문서 d 간 유사도 $S(u, d)$ 는 아래 식 (1)과 같이 상위 8개의 학생 범주 키워드 k^u 에 대해 상위 15개의 게시물 키워드 k^d 사이에 cosine 유사도(S_C)를 측정해 가장 높은 유사도의 합으로 결정한다. 이때 학생 범주 키워드에 가중치(w_i)를 부여할 수 있고, 유사도가 높은 게시물 키워드가 제목에 있을 때 가중치(w_j^T)를 부여할 수 있다.

$$S(u, d) = \sum_i^8 \max_j (w_i w_j^T S_C(k_i^u, k_j^d)) \quad (1 \leq j \leq 15) \quad (1)$$

4.5 추천 게시물 요약

전자게시판에 가독성을 고려하여 추천 게시물을 표시하기 위해서는 표시되는 게시물의 글자를 일정한 양으로 제한할 필요가 있다. 이를 위해서 자동 텍스트 요약(Automatic text summarization)을 사용할

수 있는데 이는 자연어처리 분야의 주요 연구 주제이다. 텍스트 요약에는 입력 텍스트에 있는 문장을 이용해 요약문을 구성하는 추출(Extractive) 요약과 모델이 새로운 요약문을 생성하는 추상(Abstractive) 요약이 있다. 최근 ChatGPT와 같은 대규모 언어모델에서 우수한 성능의 추상 요약을 제공하고 있지만 본 시스템에서는 추천 게시물 원문에 충실한 요약을 하는 추출 요약이 더 적절하다고 판단하여 가장 전통적인 추출 요약 알고리즘인 구글의 PageRank[24] 알고리즘에 기반한 TextRank[6] 모델을 사용하였다. TextRank 모델은 PageRank의 페이지 개념을 단어의 개념으로 바꾼 것으로 텍스트로 이루어진 글에서 특정 단어가 다른 문장과 얼마만큼 관계를 맺고 있는지 계산해서 순위를 매기고 이를 이용해 중요한 단어를 중심으로 요약문을 생성하는 방식이다.

토픽 모델링을 위한 파이썬의 Gensim 라이브러리는 TextRank 모델을 이용해 문장을 요약하는 API를 제공한다. 그런데 크롤링한 웹 문서의 다양한 특성 때문에 요약문 출력이 비정상적인 경우가 발생하여 예외 처리가 필요했다. 예를 들어 크롤링 시 HTML 코드 등이 들어있는 경우 이를 제거하도록

전처리했고, 본문의 길이가 매우 짧을 때나 부제목이 이미 존재할 때는 요약물 수행하지 않고 본문이나 부제목을 그대로 사용하도록 했다.

전자게시판에 표출할 요약된 텍스트의 글자 수를 결정하기 위해 요약문의 품질과 전자게시판 화면의 가독성을 고려하여 다양한 테스트를 수행하였고, 그 결과를 토대로 최종 50자 내외로 결정하였다.

4.6 웹 기반 추천 정보 표출

맞춤형으로 선별된 게시물 요약본을 전자게시판에 표시하기 위해 웹 기반 표출 방식을 사용하였다. 일정한 수로 추출된 교내 게시물과 교외 게시물을 전자게시판에 표출하기 위해 일정한 시간마다 화면이 자동으로 슬라이딩 되도록 하였다. 게시물에 이미지가 있는 경우와 없는 경우, 게시물이 텍스트 내용 없이 이미지로만 구성된 경우, 각각 게시물을 표시하는 형식이 달라져야 하므로 이 세 가지 경우에 대한 레이아웃 템플릿을 설계하였다. 그림 8은 이미지와 내용이 있는 게시글에 대해 설계된 레이아웃을 보여준다. 게시글에 이미지가 2개 이상 있는 경우에는 처음에 나오는 이미지만 표시하도록 하였다.

전자게시판에는 학생이 키오스크를 통해 로그인하지 않을 때도 주변 사람들을 위해 게시물을 표출해야 한다. 따라서 이때는 일반적인 대학생 범주로 추출된 키워드와 유사성이 높은 게시물을 보여주게 하였다. 학생이 로그인한 후에는 화면에 학생 맞춤형 게시물이 표시되며 오른쪽 위 영역에 로그인한 학생 정보가 표시된다. 레이아웃의 로고 영역에는 게시물의 출처를 해당 기관의 로고로 표시하도록 하였다. 맞춤 게시물 전체가 2회 슬라이딩된 후에는 해당 학생이 자동으로 로그아웃되고 화면에 표시되는 게시물도 일반 학생용 게시물로 변경된다.

V. 구현 결과 및 평가

서비스 시스템의 각 모듈은 도커 컨테이너 형태로 개발하였고 모듈 간 데이터 전송은 도커 간 소켓 통신으로 구현하였다. 구현 시스템의 시험은 내부 시험 단계와 시험 운영 단계로 진행하였다. 내부

시험 단계에서는 구현 시스템이 탑재된 개발 서버에 RFID 및 QR 코드 리더를 간이로 설치하고 로컬 모니터에서 표출된 맞춤형 게시물을 확인하며 이상 유무를 파악하였다. 각 학과 및 학년에 대한 샘플 학번 입력을 통해 해당 학생 범주에 대한 맞춤형 게시물이 잘 출력되는지도 확인하였다. 또한, 매일 크롤링이 수행되어 맞춤형 게시물의 내용이 변경되므로 내부 시험 단계 동안 변경되는 게시물에 문제가 없는지 확인하였다. 이 과정에서 발생했던 문제는 주로 매칭된 게시물의 특성으로 인해 화면에 표시되는 내용에 오류가 발생하거나 표시 내용이 깨지거나 표시되지 못한 경우로 이를 처리하기 위해 해당 코드를 수정하는 작업을 반복하였다.

시험 운영 단계에서는 서비스 시스템을 학내 가상 머신 서버에 탑재하고 학내에 설치된 전자게시판에 화면이 표출되도록 하였다. 전자게시판 우측에는 RFID 리더, QR 코드 리더 및 로그인용 PC가 탑재된 키오스크를 설치하여 실제 학생들의 이용이 가능하게 하였다. 또한 게시물이 미려하기 표시될 수 있도록 웹 디자인 업체의 협조를 받아 화면 표시를 위한 CSS(Cascading Style Sheets) 코드를 개선하였다. 시험 운영 단계에서는 갱신된 디자인으로 게시물이 표시될 때 문제점은 없는지, 실제 학생들이 로그인하고 맞춤형 게시물을 확인하는 과정에서 개선할 점이 있는지 등을 포함해 전체적인 시스템 사용성을 확인하였다. 그림 8은 시험 운영 단계에서 컴퓨터공학과 4학년 남자 재학생으로 로그인했을 때 교내 맞춤형 게시물이 표출된 예시를 보여준다.



그림 8. 시험운영 단계에서의 맞춤형 게시물 표출 예
Fig. 8. Personalized post sample during test operation phase

학내에 설치된 전자게시판은 가로세로비가 23:9인 가로가 긴 형태로 화면 공간을 좌우 2개로 나누어 좌측에 전자게시판이 설치된 도서관 소식을 표출하고 우측에 맞춤형 게시물이 표시되도록 하였다. 그림 9는 전자게시판에서 실제 표출되는 게시물의 사진이다. 우측 맞춤형 게시물은 학생이 로그인하지 않았을 때 표시되는 일반 학생용 게시물이

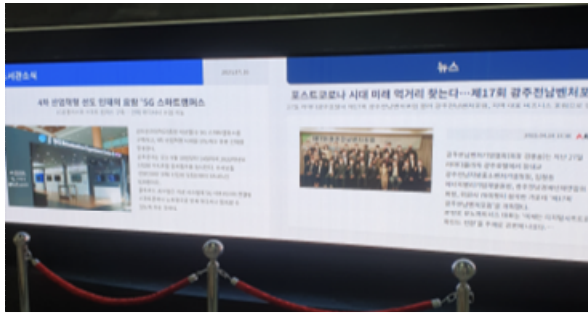


그림 9. 전자게시판에서의 맞춤형 게시물 표시 예
Fig. 9. Personalized post sample displayed on an electronic bulletin board

표 2는 제안된 시스템을 기존 관련 연구와 비교한 것이다. 제안 시스템은 스마트 전자게시판이라는 특수한 표출 환경에서의 맞춤형 추천 방식으로, 빅데이터 분석을 기반으로 한 학생 범주별 토픽모델링을 사용하였다는 점에서 기존 연구와 차별된다. 또한 단순한 키워드 매칭 방식이 아닌 워드임베딩 기반의 벡터 유사도를 사용함으로써 우수한 매칭 유사도를 기대할 수 있으며, 기사를 요약하여 표출함으로써 신속하게 기사 내용을 파악할 수 있는 장점을 가진다.

표 2. 제안 시스템과 기존 연구의 비교
Table 2. Comparison of the proposed system and existing studies

	targets	customizing method	keyword extraction	matching algorithm	summarization
this paper	news, posts	analytic topic modeling	RAKE	vector similarity	TextRank
[3]	news	view history	TF-IDF	keyword matching	-
[4]	news	view history	TF-IDF	vector similarity	-
[5]	news	view history	TF-IDF	keyword matching	-
[6]	news	view history	TF-IDF	SVD similarity	PageRank
[7]	news	manual topic modeling	TF-IDF	keyword matching	-
[8]	news	-	TF-IDF	TextRank	-

VI. 결론 및 향후 과제

제안 시스템은 학생들이 자신에게 필요한 학내의 게시물을 놓치지 않고 확인하도록 도움을 줄 수 있는 맞춤형 서비스를 제공한다. 특별히 이 시스템은 전자게시판에 연결하여 사용할 수 있도록 특화되어 있어 스마트 캠퍼스 구축 및 관련 비즈니스에도 기여할 수 있다.

제안 시스템의 게시물 추천 정확성과 유용성을 높이기 위해서는 대규모 사용 피드백을 통한 알고리즘의 고도화가 필요할 것이다. 또한 제안 시스템을 본격적인 웹이나 앱 기반의 학생 서비스로 제공하기 위해서는 대규모 크롤링과 키워드 매칭 등을 위해 많은 자원이 요구될 수 있다. 이와 관련하여 최근 본격적으로 도입되고 있는 대규모 언어 모델(LLM, Large Language Models) 기반의 챗봇 서비스와의 연계를 고려해 볼 수 있다.

References

- [1] M. Hwang, "SKT-Korea University Takes Steps Toward Implementing Metaverse-Based Smart Campus", CCTV News, Nov. 2021. <https://www.cctvnews.co.kr/news/articleView.html?idxno=231647> [accessed: Feb. 19, 2024]
- [2] Communications Team of Korea University, "'AI tutor' gets smarter, recommending extracurricular activities", Sep. 2021. https://korea.ac.kr/user/boardList.do?boardId=135&siteId=university&page=1&id=university_060102000000&boardSeq=490909&command=albumView [accessed: Feb. 19, 2024]
- [3] D. S. Park, D. H. Ahn, S. E. Hong, and H. J. Kim, "A study on recommendation method of customized articles through keyword influence analysis of news articles", In Proc. of the 2017 Summer Conference, Korean Institute of Communications and Information Science, Jeju, Korea, pp. 1537-1538, Jun. 2017.
- [4] Y.-Y. Na, J.-I. Song, and O.-R. Jeong, "A News Recommendation System Based On Vector Space

- Model", In Proc. of the Korea Computer Congress 2017, Korean Institute of Information Scientists and Engineers, Jeju, Korea, pp. 1750-1752, Jun. 2017.
- [5] H.-K. Yeo, J.-I. Song, and O.-R. Jeong, "Personalized News Recommendation Using Temporal Database", In Proc. of the Korea Computer Congress 2017, Korean Institute of Information Scientists and Engineers, Jeju, Korea, pp. 1706-1708, Jun. 2017.
- [6] I.-J. Jeong, B.-M. Kim, S.-K. Kim, and K. Yu, "News Recommendation System Based on Text and Image Tag Data", Journal of Digital Contents Society, Vol. 21, No. 3, pp. 479-486, Mar. 2020. <http://dx.doi.org/10.9728/dcs.2020.21.3.479>.
- [7] D.-H. Kim, S.-Y. Yun, S.-Y. Chae, and S.-J. Yu, "A News Recommendation System using Text Mining and LOD Technique", Journal of KIIT, Vol. 20, No. 8, pp. 1-6, Aug. 2022. <http://dx.doi.org/10.14801/jkiit.2022.20.8.1>.
- [8] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts", In Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, pp. 404-411, Jul. 2004.
- [9] K. Sin, D. Kim, and S. Choi, "A study on how to recommend news of interest to users in the power generation fuel field using user-defined topic modeling", In Proc. of the 2021 Summer Conference, The Korean Institute of Electrical Engineers (KIEE), Pyeongchang, Korea, pp. 2229-2240, Jul. 2021.
- [10] KBS, "KBS News", <https://news.kbs.co.kr/news/pc/main/main.html> [accessed: Feb. 19, 2024]
- [11] Electronic Times Internet, "Electronic Times", <https://www.etnews.com> [accessed: Feb. 19, 2024]
- [12] AI Times, "AI Times", <https://www.aitimes.com> [accessed: Feb. 19, 2024]
- [13] JOBKOREA LLC, "JOBKOREA", <https://www.jobkorea.co.kr> [accessed: Feb. 19, 2024]
- [14] THINKGOOD, "Thinkgood", <https://www.thinkcontest.com> [accessed: Feb. 19, 2024]
- [15] NAVER Corp., "Naver News", <https://news.naver.com> [accessed: Feb. 19, 2024]
- [16] KOREA PRESS FOUNDATION, "BIG Kinds", <https://www.bigkinds.or.kr> [accessed: Feb. 19, 2024]
- [17] NAVER Corp., "DataLab", <https://datalab.naver.com> [accessed: Feb. 19, 2024]
- [18] Kumoh National Institute of Technology, "KIT Integrated Search", <https://www.kumoh.ac.kr/srch/index.do>, [accessed: Feb. 19, 2024]
- [19] National Information Society Agency, "Integrated Search", https://www.nia.or.kr/site/nia_kor/ex/search/searchList.do [accessed: Feb. 19, 2024]
- [20] RAKE, "A Python library that can extract keywords from any document or a piece of text", <https://github.com/u-prashant/RAKE> [accessed: Feb. 19, 2024]
- [21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space", arXiv: 1301.3781v3 [cs.CL], Jan. 2013. <https://doi.org/10.48550/arXiv.1301.3781>.
- [22] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information", arXiv:1607.04606v2 [cs.CL], Jul. 2017. <https://doi.org/10.48550/arXiv.1607.04606>.
- [23] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation", Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1532-1543, Oct. 2014. <https://doi.org/10.3115/v1/D14-1162>.
- [24] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking : Bringing Order to the Web", The Web Conference, pp. 161-172, Nov. 1999.

저자소개

김 태 형 (Tae-Hyong Kim)



1995년 8월 : 연세대학교
전기전자공학과(공학석사)
2001년 2월 : 연세대학교
전기전자공학과(공학박사)
2002년 9월 ~ 현재 :
금오공과대학교 컴퓨터공학과,
인공지능공학과 교수

관심분야 : 머신러닝, 데이터분석, 스마트팩토리

김 민 영 (Min-Young Kim)



2023년 2월 : 금오공과대학교
컴퓨터공학과(공학사)
관심분야 : 머신러닝, 데이터분석

김 주 엽 (Ju-Yeop Kim)



2023년 2월 : 금오공과대학교
컴퓨터공학과(공학사)
관심분야 : 클라우드 SA

전 지 용 (Ji-Yong Jeon)



2023년 8월 : 금오공과대학교
컴퓨터공학과(공학사)
관심분야 : 머신비전, 멀티모달

정 은 혜 (Eun-Hye Jeong)



2023년 2월 : 금오공과대학교
컴퓨터공학과(공학사)
관심분야 : 머신러닝, 풀스택

현 소 담 (So-Dam Hyeon)



2023년 8월 : 금오공과대학교
컴퓨터공학과(공학사)
관심분야 : MLOps