

**Python Programming and Practice**

# **Development daily news keyword summary program**

**Final Report**

Date : 2023.12.24

Name : 지연우

ID : 224491

## **1. Introduction**

### **1) Background**

It's time consuming and very cumbersome to check out all the tons of news pouring out every day. But we can't completely stop paying attention to current events, so we need a more efficient way to get information. To solve this problem, a program that can easily check current events information in short sentences or keywords is needed.

### **2) Project goal**

Aim to create a program that summarizes articles uploaded every day in short sentences and a few words.

### **3) Differences from existing programs**

The article summary function previously provided by Naver News only summarizes each article. This, of course, helps save time, but it is difficult to easily find out various current events information. We analyze all articles to show the user a few key sentences or keywords, making it easy to get current events information.

## **2. Functional Requirement**

### **1) Collect articles**

- function to collect articles uploaded daily

#### **(1) Collect information on popular articles by media company**

- Crawl the titles, body text of the top five popular articles of each media company

### **2) Summarize the articles**

- Summarize each article in short sentences and words.

#### **(1) Create article summary**

- Summarize the article in one sentence.

#### **(2) Extract keywords**

- Extract one keyword of articles.

### **3) Show keywords**

- Show keywords to user.

#### **(1) Sort keywords in order of frequency**

- Show the keywords that appear in more articles first.

### **4) Show summary of articles by keyword**

- When a user selects a keyword, it shows a summary of the article in which that keyword appears. And when a user enters article number, it shows article body text.

## 3. Implementation

### (1) Collect articles

- Input: Link to Naver News Ranking page
- Output: TSV file containing information of the articles
- Description: Crawl the titles, body text and url of the top five popular articles from each media company.
- Applied concepts: loops, functions, string manipulation, exception handling, file write, `__name__ == "__main__"`, package, web crawling
- Code Screenshot

```
from datetime import datetime
import os.path

import requests
from bs4 import BeautifulSoup as bs

# 주어진 링크의 html파일을 lxml로 파싱한 BeautifulSoup 객체를 반환하는 함수
2 usages  ↳ Yeonwoo Ji
def get_soup(link):
    USER_AGENT = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/119.0.0.0 Safari/537.36"
    head = {'user-agent': USER_AGENT}
    req = requests.get(link, headers=head)
    html = req.text
    soup = bs(html, features="lxml")
    return soup

# 링크가 주어진 기사의 정보를 리스트로 만들어 반환하는 함수
# [기사제목, 기사링크, 기사분류] 형태로 반환
1 usage  ↳ Yeonwoo Ji +1
def crawl_article(article_link):
    soup = get_soup(article_link)

    title = soup.select_one("#title_area > span").text
    body_text = soup.select_one("#dic_area").text.strip().replace(_old: "\n", _new: " ").replace(_old: '\t', _new: ' ')

    return [title, article_link, body_text]

# 네이버 뉴스의 월경뉴스 페이지에서 언론사별 조회수 상위 5개 기사의 정보를 수집하는 함수
# 각 기사에 대해 crawl_article 함수를 사용하여 기사 정보를 얻어오고, tsv파일에 저장
3 usages  ↳ Yeonwoo Ji +1
def crawl_rankings(file_name):
    if os.path.isfile(file_name):
        return
    rankings_link = "https://news.naver.com/main/ranking/popularDay.naver"
    soup = get_soup(rankings_link)
    rankingnews_boxes = soup.select("#rankingnews_box") # 언론사별 월경 5위까지의 기사 정보가 담긴 박스들 선택

    with open(file_name, 'w', encoding="utf8") as fp:
        fp.write("언론사명\t기사제목\t기사링크\t기사분류\n")

        for box in rankingnews_boxes:
            company_name = box.select_one("#rankingnews_name").text
            # 영어 기사는 제외하도록 영문 신문은 수집하지 않음
            if company_name in ("코리아헤럴드", "코리아중앙데일리"):
                continue
            articles = box.select("li")

            for i, article in enumerate(articles):
                try:
                    # 링크는 https://n.news.naver.com/article/언론사코드/기사번호?ntype=RANKGIN 형태
                    article_link = article.select_one(".list_title")["href"]
                    article_data = crawl_article(article_link)
                    fp.write(company_name + "\t" + "\t".join(article_data) + "\n")
                    # 집계기준에 해당하는 기사가 없어 월경 5위까지 기사가 존재하지 않을 때
                except:
                    print(company_name, i+1, "변태 기사 수집 실패")
```

## (2) Summarize the articles

- Input: Crawled articles data
- Output: Save TSV file containing information of the articles, summaries of the articles and keywords of the articles. And Return pandas dataframe of that file and a dictionary with a keyword extracted from the articles as a key and a list of indexes of the sentence in which the keyword appeared as a value
- Description: Summarizes articles using t5 models and calculates the frequency of keyword appearance by extracting keywords extracted by keybert.
- Applied concepts: loops, functions, list comprehension, file read, file write, dictionary, pandas, keyword extraction, text summarize, package
- Code Screenshot

```
import os.path

import nltk

from transformers import AutoTokenizer, AutoModelForSeq2SeqLM
from konlpy.tag import Hannanum
from keybert import KeyBERT
import pandas as pd

# https://huggingface.co/eenzeenee/t5-base-korean-summazation t5기반 한국어 문서 요약 모델을 사용하여 기사들의 요약문 추출
# 기사 10개씩 나누어서 요약하도록 설정
!usage -s speciling
def get_summarize(articles):
    nltk.download('punkt')

    model = AutoModelForSeq2SeqLM.from_pretrained('eenzeenee/t5-base-korean-summazation').to('cuda')
    tokenizer = AutoTokenizer.from_pretrained('eenzeenee/t5-base-korean-summazation')
    result = []

    for i in range(0, len(articles), 30):
        inputs = ["summarize: " + article for article in articles[i:min(i+30, len(articles))]]
        inputs = tokenizer(inputs, max_length=512, truncation=True, return_tensors='pt', padding=True).to('cuda')
        output = model.generate(**inputs, num_beams=3, do_sample=True, min_length=10, max_length=100)
        decoded_output = tokenizer.batch_decode(output, skip_special_tokens=True)
        result += decoded_output

    return result

# 요약문에서 불용어가 아닌 명사만을 추출하여 키워드로 사용. keyword_dic에 {키워드: 포함하는 문장 번호의 리스트} 형식으로 데이터 저장
!usage -s speciling
def get_keywords(articles):
    #key_model = KeyBERT()
    key_model = KeyBERT('paraphrase-multilingual-MiniLM-L12-v2')
    hannanum = Hannanum()
    keyword_dic = {}
    stop_words = {'등', '씨', '것', '여성', '혐의', '경찰', '서울', '한국', '긴급', '불쾌', '이탈리', '10대', '20대', '30대',
                  '40대', '50대', '60대', '70대', '80대', '분노', '한', '징역', '재', '남성', '사고', '평균', '대한민국'}
    keywords = [key_model.extract_keywords(article, keyphrase_ngram_range=(1,2), top_n=2) for article in articles]
    keywords = [k[0][0] + ' ' + k[1][0] for k in keywords]
    for i, summary in enumerate(keywords):
        words = set([word for word in hannanum.nouns(summary) if word not in stop_words])
        keywords[i] = words
        for word in words:
            if word in keyword_dic:
                keyword_dic[word].append(i)
            else:
                keyword_dic[word] = [i]
    return keywords, keyword_dic
```

```
def summarize_article(file_name, processed_file):
    if os.path.isfile(processed_file):
        df = pd.read_csv(processed_file, sep='\t')
        keyword_dic = {}
        for i, words in enumerate(df['키워드']):
            for word in set(words[1:-1].split(',')):
                word = word.strip().strip('"')
                if word in keyword_dic:
                    keyword_dic[word].append(i)
                else:
                    keyword_dic[word] = [i]
        return df, keyword_dic
    df = pd.read_csv(file_name, sep='\t')
    df['기사요약'] = get_summarize(df['기사본문'])
    df['키워드'], keywords_dict = get_keywords(df['기사본문'])
    df.to_csv(processed_file, index=False, sep="\t")
    return df, keywords_dict
```

### (3) Show keywords

- Input: processed article data
- Output: keywords sorted in order that appear in more articles
- Description: I defined a CLI class that organizes the screen for use in the CLI environment. The function to show keywords was implemented as a member function called `print_keywords` inside the class. This function sorts keywords in the order that they appear in more articles and shows them to the user 10 at a time. Users can move on to the next page or the previous page by typing on the keyboard, and each time they do so, the screen is automatically erased and redrawn. When a user selects a keyword, the `print_titles` function is called to view articles featuring the keyword.
- Applied concepts: class, loops, functions, f-string, dictionary, pandas, package
- Code Screenshot

```

class CLI():
    # specilling
    def __init__(self, df, keywords_dict):
        self.df = df
        self.keywords_dict = keywords_dict
        self.system = platform.system()

    # specilling
    def run(self):
        self.print_keywords()

1 usage # specilling
    def print_keywords(self):
        keywords = sorted(self.keywords_dict.keys(), key=lambda x: -len(self.keywords_dict[x]))
        keywords_set = set(keywords)
        n = 0 # n번부터 10개의 키워드를 보여줌
        while True:
            self.clear()
            print("-----")
            print("-----      현재 기사에서 많이 언급되는 키워드      -----")
            print("-----")
            print("-----      이전 10개 보기: b      -----")
            print("-----      다음 10개 보기: n      -----")
            print("-----      키워드가 등장한 기사 보기: 번호 or 키워드      -----")
            print("-----      프로그램 종료: exit or ctrl+c      -----")
            print("-----")
            print()
            for i, keyword in enumerate(keywords[n:n + 10]):
                print(f'{n + i}. {keyword}')
            print()
            while True:
                try:
                    command = input(">>")
                except:
                    break
                if command == "b":
                    if n - 10 >= 0:
                        n -= 10
                        break
                    else:
                        print("첫 페이지입니다. 다른 명령어를 입력해주세요.")
                        continue
                elif command == "n":
                    if n + 10 < len(keywords) - 1:
                        n += 10
                        break
                    else:
                        print("마지막 페이지입니다. 다른 명령어를 입력해주세요.")
                        continue
                elif command == "exit" or command == "quit":
                    return
                elif command.isdecimal() and (keyword_num := int(command)) < len(keywords):
                    self.print_titles(keywords[keyword_num])
                    break
                elif command.isalnum() and command in keywords_set:
                    self.print_titles(command)
                    break
                else:
                    print("올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.")

```

#### **(4) Show summary of articles by keyword**

- Input: processed article data, selected keyword
- Output: information on articles featuring keywords and the body of articles selected by the user
- Description: The function to show summary of articles by keyword was implemented as a member function called `print_keywords` inside the `CLI` class. This function shows five pieces of information (article title, media company name, article summary) in an article containing the keywords you choose. Users can move on to the next page or the previous page by typing on the keyboard, and each time they do so, the screen is automatically erased and redrawn. When a user enters the article number that they want to take a closer look at, it shows the article's title, link, body and keywords in detail.
- Applied concepts: class, loops, functions, f-string, dictionary, pandas, package
- Code Screenshot



```

def print_titles(self, keyword):
    n = 0 # n번부터 5개의 기사 정보(제목, 언론사, 요약문)를 출력해줌
    article_nums = self.keywords_dict[keyword]
    while True:
        self.clear()
        print("-----")
        print(f"----- 키워드 '{keyword}' 가 언급된 기사 -----")
        print("-----")
        print(f"----- 이전 5개 보기: b -----")
        print(f"----- 다음 5개 보기: n -----")
        print(f"----- 기사 전문 보기: 기사 번호 입력 -----")
        print(f"----- 키워드 보기로 돌아가기: exit -----")
        print("-----")
        print()
        for article_num in article_nums[n:n + 5]:
            article = self.df.loc[article_num]
            print(f"{article_num}번 기사")
            print(f"{article['기사제목']}\t\t{article['언론사명']}")
            print(f"{article['기사요약']}")
            print()
        while True:
            try:
                command = input(">>")
            except:
                break
            if command == "b":
                if n - 5 >= 0:
                    n -= 5
                    break
                else:
                    print("첫 페이지입니다. 다른 명령어를 입력해주세요.")
                    continue
            elif command == "n":
                if n + 5 < len(article_nums) - 1:
                    n += 5
                    break
                else:
                    print("마지막 페이지입니다. 다른 명령어를 입력해주세요.")
                    continue
            elif command == "exit" or command == "quit":
                return
            elif command.isdecimal() and (article_num := int(command)) in article_nums:
                self.clear()
                article = self.df.loc[article_num]
                print(f'기사제목\t{article["기사제목"]}\n')
                print(f'언론사명\t{article["언론사명"]}\n')
                print(f'기사링크\t{article["기사링크"]}\n')
                article_text = article["기사본문"].replace(".", ".\n") # 가독성을 위해 마침표 뒤에 줄바꿈 추가
                print(f'기사본문\n{article_text}\n')
                print(f'키워드\t{article["키워드"]}\n')
                input("엔터를 눌러 나가기>>")
                break
            else:
                print("올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.")

3 usages  ▲ speciling
def clear(self): # os별 터미널 화면 초기화 명령어 사용, 코랩에서는 정상적으로 작동하지 않음.
    if self.system == 'Windows':
        os.system("cls")
    else:
        os.system("clear")

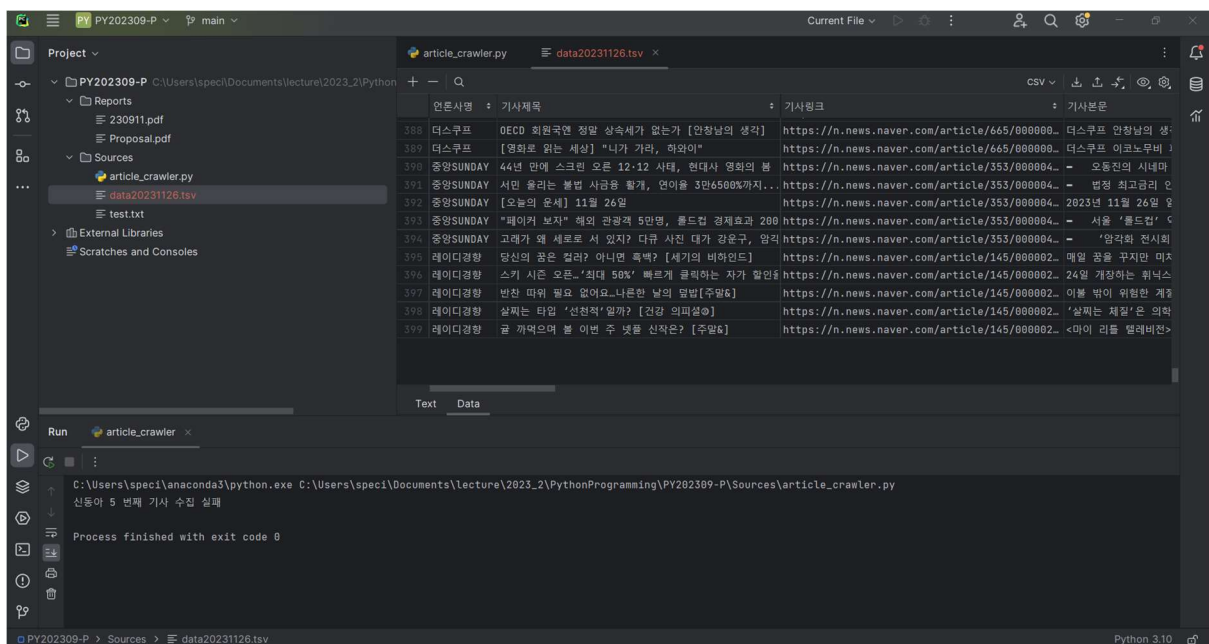
```

## 4. Test Result

### (1) Collect articles

- Description: Successfully crawled information from 399 articles across 80 media outlets on Naver News Ranking page. For 'Shindonga' (a specific media outlet), only articles up to the 4th rank were available, resulting in a failure message when attempting to collect the 5th article.

- Test Results Screenshot:



### (2) Summarize the articles

- Description: The article summary worked very well, and keyword extraction also performed not bad using keybert.

- Test Results Screenshot:

----- 기사제목: 요약문 -----

"거의 다 먹었는데..."스키장 식당 급식 밥통서 나온 회색 수건: 강원도 한 스키장 식당의 단체급식용 밥통에서 사용하던 수건이 나와 이용객이 눈살을 찌푸리는 일이 발생했다.

'K배터리' 큰일났다...'승률 80%' 트럼프 돌아오면 벌어질 일: 트럼프 전 미국 대통령이 미국 대선에서 재집권에 성공하면 보복 관세를 확대하고 친환경 에너지 지원책을 대거 폐기하겠다고 선언했다.

[단독] 믿었던 대학병원의 배신...'가짜 아킬레스건 수술' 뒤통수: 한국경제신문이 확보한 무허가 인체조직 이식 현황을 분석한 결과 전국 232개 병원에 반쪽 아킬레스건 수술이 유통된 것으로 확인됐으며 전문 아집밥 정겨주코 오후 3시 칼퇴' 출산을 3배' 기적 일어났다: 이트주상사의 사내 할계특수출산율은 2021년 1.97명으로 뛰어 10년 만에 일본 3대 종합상사의 출산율이 세 배 넘게 된 현상을 일본 사회도 기

"Y6 주식 미리 사들일 걸?"주가 폭등하자 후회 쏟아졌다 [연계소문]: Y6연티테인트가 그룹 블랙핑크와 완전히 활동에 대한 재계약을 체결했다고 밝히자 업계 안팎에서 후회의 목소리가 나왔다.

"남성 둘과 중거, 키 크고 예뻐다"...35년전 조지아주 피살 한인여성 신원 확인: SBS 그것이 알고 싶다가 35년 동안 미제로 남아있던 김정은 씨 피살사건의 전모를 파헤쳤다. 한적한 마을의 쓰레기 수거함에서

'별거아내와 조대남' 성관계 영상 SNS 올린 군 간부 징유...왜: 의정부지법 형사3단독은 SNS에 성관계 영상을 올린 혐의로 기소된 전방지역 직업군인 A 씨에게 징역 1년에 집행유예 2년을 선고했다.

수백 명 먹었는데...스키장서 밥 툄다 따라 나온 수건 논란: 강원도 내 한 스키장의 모식당에서 식사를 한 고객이 급식용 밥통에서 수건을 발견해 논란이 되자 해당 업체는 같은 사례가 발생하지 않도록 조치하

이준석 "신당, 1월 안에 매듭...이낙연, 같이 그리는 울타리 돼야": 이 전 국민의힘 대표는 윤 대통령의 변화를 유일한 변수라고 말하며 신당 창당과 관련해 늦어도 1월 안에 매듭지을 수 있을 것이라고 밝혔다

수능 수학 1등급 96.5%가 이과생...올해도 '문과집중' 여전: 2024학년도 대학수학능력시험에서 수학 1등급 학생 중 96.5 가 미적분을 선택한 이과 학생인 것으로 나타났으며 과목별 유희리가 두드러지면서 국

'해군 광파'였던 55세 여성이 '오징어게임' 실사판' 상급 60억 따낸 비결은: 넷플릭스 오징어게임 더 캘린지 우승자인 최한은 서바이벌 쇼에 출연한 뒤 두려움 없이 자신이 된 삶을 해쳐 나가겠다고 세계 시청

수험생 흥분 거저 광고, 업체 1~4위 학원 다 걸렸다: 입시업계 최대 학원인 메가스터디와 시대인재 등이 대학수학능력시험 수험생을 끌어들이기 위해 부당 광고를 한 혐의로 과징금 재정을 맞았다. 이들은 소

국민의힘 영입인재 1호 "이 당이나 저 당이나 모두 퇴보 정치" 쓴소리: 국민의힘이 인제 1호로 영입한 하 소아청소년과외원 원장은 육아 방법에 대해 조언하고 있다. 유튜브 캡처 하 소아청소년과외원 원장이

내일 출근길부터 전국에 많은 비... 강원도는 대설 주의: 중국 내륙에서 발달한 저기압으로 인해 많은 양의 비가 내리면서 일부 지역에는 호우특보가 내려질 수도 있고 다음 주말부터는 한파가 찾아올 전망이다

"수능 수학 1등급 96.5%가 이과생"... '문과 집중' 한층 거세진다: 2024학년도 수능에서 수학 영역 1등급의 이과생 독자지 현상이 심화됐다는 입시업계 분석이 나왔다.

[단독] "SPC 불매 탓 매출 70% 급락" 단전 점포 두 곳 손배소: 에스피씨그룹 산하 빌알코리아 계열 단전 점주들은 본사의 위법행위로 매출이 70 이상 급락해 수익원의 손해를 입었다고 주장하며 소송을 제

서울의 봄 전두환 '김식당' 열장은 실화...쿠팡타 자축하며 한 말은: 영화 서울의 봄이 흥행을 이어가며 12 12군사반란과 실제 역사를 비교하며 영화를 즐기는 관객들도 늘고 있다. 특히 영화는 극단의 긴장으

'사교육 카르텔' 저축 학원서 만점·표점 최고점...공립들 '킬러 배제': 2024학년도 대학수학능력시험에서 수학 영역 1등급 수험생 96 가량이 미적분이나 기하에 응시한 이과생인 것으로 나타나 정부가 수능에서

'원칙과 상식', 국회서 세 과시...창당 뜻 내비치며 지도부 압박: 더불어민주당 내 비주류 의원모임인 원칙과 상식이 대우형 행사를 열어 당 지도부 압박에 나섰다.

새벽일 못 시켜 해고, 워킹맘이 이었다...'회사-육아지원' 첫 관례: 어린 자녀를 키우는 노동자에게 새벽 공휴일 근무를 강요하고 이를 받아들이지 않자 채용을 거부하는 것은 부당해고라는 대법원 판단이 나왔

cli.py processedData20231224.tsv		processedData20231224.tsv	
C5		C6	
기사요약		키워드	
스타트업 AI 동영상 생성 서비스 애플리케이션인 피카 1.0을 출시한 피카랩스는 자연스러운 움직임과 사실적인 영상 품질 등도		{ '1995년생', '창업', '스타트업' }	
키트가 전 법무부 장관의 딸 조 씨가 유전자 검사 결과를 공개하면서 DNA 혈통 검사에 대한 관심이 높아졌다.		{ '검사', '혈통분석', '유전자' }	
석공연: 유적집 업체에 고착화된 관망 형태를 지칭하는 시체관공 논란이 일고 있는데 관객들의 웃자림이 두꺼워지자 공연 시		{ '유적집', '업체', '대중음악계' }	
보니상성 개인들의 삼성전자 순매도세가 거세다. 시장에서 2차전지에서 반도체로 주도주가 바뀌고 있지만 개인들의 매수는 2		{ '소액주주', '나타나면서다', '삼성전자' }	
전기안전 세종시의 한 목욕탕에서 70대 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌으나 모두 숨졌다.		{ '세종충남대병원', '목욕탕', '세종시' }	
격을 두. 남성 버스 기사가 버스 안에서 홀로 있는 여성 승객을 두고 종이컵에 소변을 봤다. c News1 김 디자이너 남성 버스		{ '버스', '버스기사' }	
파환파 크리스마스 연휴가 시작된 가운데 대전 성심당 인근 지역이 케이크 구매를 위한 인파로 장사진을 이뤘다는 후기가 올		{ '크리스마스' }	
장년39대 세종 조치원읍의 한 목욕탕에서 감전사고가 발생해 70대 여성 3명이 숨졌는데 소방당국은 심정지 상태인 이들을 긴급		{ '목욕탕', '감전사고' }	
수립 포: 조 전 법무부 장관은 자신의 페이스북을 통해 원복한 검찰공화국의 수립을 위한 포석이 높였으며 검찰 출신에 장악된		{ '장관', '법무부', '낙점', '비상대책위' }	
출소 12 여신도들을 성폭행해 1심에서 중형을 선고받은 기독교복음선교회 정 종재가 대전지법 형사12부에 항소장을 제출했고		{ '기독교선복음교회', '기독교복음선교회' }	
[단독] 신 정 전세계그룹 부회장은 아내 한 씨의 독주회에 참석해 대기실에서 손님을 맞았고 취재진과 관객들에게 아들을 직접		{ '부회장', '취재진', '성탄절' }	
배우 이.마약 투약 혐의를 받는 배우 이 씨가 마약 관리에 관한 법률상 대마 및 향정 현의를 받는 이 씨를 피의자 신분으로		{ '마약범죄수사계', '인천경찰청' }	
3세인 유적집 배우 함 씨가 유튜브 채널 현연지에서 올린 영상에서 업로드를 중단한다고 발표했다.		{ '현연지', '음식', '유튜브' }	
서울경쳐 남성 버스 기사가 버스 안에서 홀로 있는 여성 승객을 두고 종이컵에 소변을 보는 일이 있었는데 여성은 경찰에게 도		{ '버스', '탐송한' }	
] 여성: 이스라엘 바이츠만연구소의 소벨 신경생물학 교수 연구팀이 발표한 연구 결과에 따르면 타인의 눈물 냄새를 맡은 남		{ '생물학', '연구팀' }	
야 패션 이탈리아 패션 인플루언서가 소아암 어린이를 위해 기부된다며 비싼 크리스마스 케이크를 홍보해 15억 원이 넘는 벌		{ '케이크', '반독점' }	
목욕탕에 세종시의 한 목욕탕에서 70대 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌으나 모두 숨졌고 소방 당국: { '목욕탕' }		{ '목욕탕' }	
실장은 5 이 대통령실 정책실장은 더불어민주당이 강행 처리를 추진 중인 특별검사 도입 법안들에 대해 총선을 겨냥해 흡집내		{ '오늘', '정책실장' }	

### (3) Show keywords

- Description: It was confirmed that 10 keywords appeared well in the CLI environment, and the page was moved well. If a user enters a keyword number or keyword, information on articles containing keywords will appear normally. It also works well when a user enters incorrect commands, missing keyword numbers, or missing keywords.

- Test Results Screenshot:

```

-----
현재 기사에서 많이 언급되는 키워드
-----
이전 10개 보기: b
다음 10개 보기: n
키워드가 등장한 기사 보기: 번호 or 키워드
프로그램 종료: exit or ctrl+c
-----

0. 목욕탕
1. 국민의힘
2. 크리스마스
3. 장관실장
4. 정책실장
5. 삼성전자
6. 마약범죄수사계
7. 케이크브
8. 유튜거
9. 냉장고

>>b
첫 페이지입니다. 다른 명령어를 입력해주세요.
>>

```

```

-----
현재 기사에서 많이 언급되는 키워드
-----
이전 10개 보기: b
다음 10개 보기: n
키워드가 등장한 기사 보기: 번호 or 키워드
프로그램 종료: exit or ctrl+c
-----

760. 축복
761. 파트1
762. 넷플릭스
763. 성
764. 결혼
765. 파티
766. 기분
767. 아이돌
768. 길
769. 생명력

>>n
마지막 페이지입니다. 다른 명령어를 입력해주세요.
>>n
마지막 페이지입니다. 다른 명령어를 입력해주세요.
>>?
올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.
>>999
올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.
>>없는키워드
올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.
>>

```



## (4) Show summary of articles by keyword

- Description: Up to five pieces of article information are well displayed at once, and it was confirmed that the detailed information of the article was well displayed when the article number was entered. Like the above function, processing of incorrect user input works well.
- Test Results Screenshot:

```
Anaconda Prompt - python main.py

----- 키워드 '목욕탕' 가 언급된 기사 -----
-----
----- 이전 5개 보기: b -----
----- 다음 5개 보기: n -----
----- 기사 전문 보기: 기사 번호 입력 -----
----- 키워드 보기로 돌아가기: exit -----

4번 기사
탕 안에서 비명 지르며 쓰러졌다...세종시 목욕탕 3명 '참변' [종합] 한국경제
세종시의 한 목욕탕에서 70대 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌으나 모두 숨졌다.

7번 기사
온탕서 '악'하며 쓰러져...세종 목욕탕 감전 3명 모두 사망(종합2보) 뉴스1
세종 조치원읍의 한 목욕탕에서 감전사고가 발생해 70대 여성 3명이 숨졌는데 소방당국은 심정지 상태인 이들을 긴급 이송했고 이 가운데 2명은 병원으로 이송하던 중 숨졌다.

16번 기사
세종시 목욕탕에서 감전사고 벌어져 70대 3명 감전사 SBS
세종시의 한 목욕탕에서 70대 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌으나 모두 숨졌고 소방 당국과 전기안전공사 등은 이들이 감전된 것으로 보고 있다.

20번 기사
세종시 목욕탕서 여성 입욕객들 감전...3명 사망 KBS
세종시의 한 목욕탕에서 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌지만 3명 모두 숨졌고 경찰과 소방 당국은 정확한 사고 원인을 조사하고 있다.

41번 기사
목욕탕 감전으로 의식 없던 70대 숨져...사망자 3명으로 늘어 뉴시스
세종시 조치원읍에서 발생한 목욕탕 전기 감전 사고로 의식이 없던 70대 여성 1명이 끝내 숨졌고 소방본부에 따르면 목욕탕 여탕에서 악 소리를 들은 목욕자가 신고했다.

>>b
첫 페이지입니다. 다른 명령어를 입력해주세요.
>>
```

```
Anaconda Prompt - python main.py

----- 키워드 '목욕탕' 가 언급된 기사 -----
-----
----- 이전 5개 보기: b -----
----- 다음 5개 보기: n -----
----- 기사 전문 보기: 기사 번호 입력 -----
----- 키워드 보기로 돌아가기: exit -----

264번 기사
세종시 목욕탕서 감전 사고...입욕객 3명 모두 사망 채널A
세종시 조치원읍 한 목욕탕에서 감전사고가 발생해 70대 여성 3명이 숨졌는데 목욕객 등의 진술로 봤을 때 감전 가능성이 높다.

286번 기사
세종 목욕탕 감전으로 70대 여성 3명 모두 사망 여성신문
세종시 조치원읍에서 발생한 목욕탕 전기 감전 사고로 의식이 없던 70대 여성 1명이 끝내 숨졌으며 소방 당국과 전기안전공사 등은 이들이 감전된 것으로 보고 있다.

296번 기사
세종시 '목욕탕 참변' ...결국 3명 모두 사망 농민신문
세종시 재난안전대책본부 등에 따르면 세종시에 있는 한 목욕탕에서 70대 여성 이용객 3명이 감전돼 심정지 상태로 병원으로 이송됐으나 결국 3명 모두 사망했다.

368번 기사
온탕 들어갔다 '악'...세종서 70대 여성입욕객 3명 모두 사망 대전일보
세종시 조치원읍에 위치한 목욕탕에서 70대 여성 입욕객 3명이 감전 및 심정지 상태로 병원으로 옮겨졌으나 모두 숨졌다.

>>n
마지막 페이지입니다. 다른 명령어를 입력해주세요.
>>300
올바르지 않은 명령어입니다. 명령어를 다시 확인해주세요.
>>
```

```
Anaconda Prompt - python main.py

기사제목 탕 안에서 비명 지르며 쓰러졌다...세종시 목욕탕 3명 '참변' [종합]
언론사명 한국경제
기사링크 https://n.news.naver.com/article/015/0004828782?type=RANKING
기사본문
세종시 목욕탕서 70대 여성 3명 감전사 지난 6월 전기안전점검서는 '이상 없음'결함 "온수탕은 전기 출려 들어간 등...정밀 감시 중" 사진은 기사와 직접적인 관계 없음.
기타이미지제공24일 새벽 세종시의 한 목욕탕에서 70대 여성 입욕객 3명이 감전돼 심정지 상태로 병원으로 옮겨졌으나 결국 모두 숨졌다.
세종시와 세종경찰청 등에 따르면 이날 오전 5시 37분경 조치원읍 목욕탕의 온탕 건물 지하 1층 여성 목욕탕에서 온수탕 안에 있던 A·B(이상 71)·C(70)씨가 비명을 지르며 쓰러지는 것을 목격자 등이 보고 119에 신고했다.
당시 목욕탕에는 목욕객이 다 있었으나, 온수탕에 들어간 뒤 3명만 병을 입었다.
경찰에 출동한 소방 당국은 심정지 상태인 이들을 중환자병원에 '심정지나병행' '세종충남대병원'으로 긴급 이송했으나, 끝내 모두 숨졌다.
당국과 소방 당국은 '온수탕 침몰' 등 이들이 감전된 것으로 보고 있다.
경찰 관계자는 "온수탕 침몰로 전기기가 물에 들어간 것으로 보인다"며 "누전 원인을 밝히기 위해 오후 1시부터 합동 감시를 벌이고 있다"고 말했다.
지난 6월 이 목욕탕에 대한 전기안전공사 안전 점검 때는 별다른 문제가 발견되지 않은 것으로 알려졌다.
사고가 난 건물은 지하 1층 지상 3층 규모로, 전라 전이 1994년 12월 사용 승인했다.
지하 1층은 여탕(173㎡)과 보양실(99㎡), 지상 1층은 카운터와 남탕, 2~3층은 도발로 사용돼 왔다.
세종시는 이번 사고를 계기로 지역 내 목욕탕 20여곳의 전기안전점검 철저, 점검할 예정이다.
2018년 10월 29일 새벽 3명의 한 사우나 탕 안에서 3명 중이던 70대 여성 3명이 감전돼 숨지고 여탕에 있던 2명도 다쳤는데, 당시 사고에 목포수출 공급하려고 설치한 모래에 연결돼 있던 전선이 끊어져 발생한 것으로 결론 났다.
이 사우나 업주와 전기안전관리자 등은 업무상 과실치사로 벌금 400만~700만원 약식명령을 받았다.

키워드 {'세종충남대병원', '목욕탕', '세종시'}
엔터를 눌러 나가기>>
```

## **5. Changes in Comparison to the Plan**

### **1) Removal of Comment Count-related Feature**

- Previously: Comment count was collected in Function 1, and when showing article summaries in Function 4, they were displayed in descending order of comment count.
- Now: Comment count is no longer collected in Function 1. In Function 4, article summaries are displayed randomly.
- Reason: Comment count information was not present in the crawled HTML documents. Although considering using Selenium to collect comment counts was an option, it was decided to exclude this feature as it was not deemed crucial, and the time required for article collection became excessively long.

## **6. Lessons Learned & Feedback**

I brought an external model and tried it for the first time, and I think it was a great experience to learn how to use the trained model. Thank you for your passionate instruction during the semester!