

“ 청년인턴

데이터 분석

우수사례

”

Best Practice

우수사례 5. 빅데이터 분석을 활용한 공직감찰 수요분석

“**먼저 찾는 비판기사,**
먼저 푸는 감찰 대응!”

Index

비판기사 수집모델



개요

프로젝트 필요성, 목표



데이터 정제

검색 키워드, 신문사 선정



비판기사 수집 GUI 제작

크롤러 생성, GUI 제작



활용 및 기대효과

GUI 프로그램 시연, 활용방안



프로젝트 개요

프로젝트 필요성

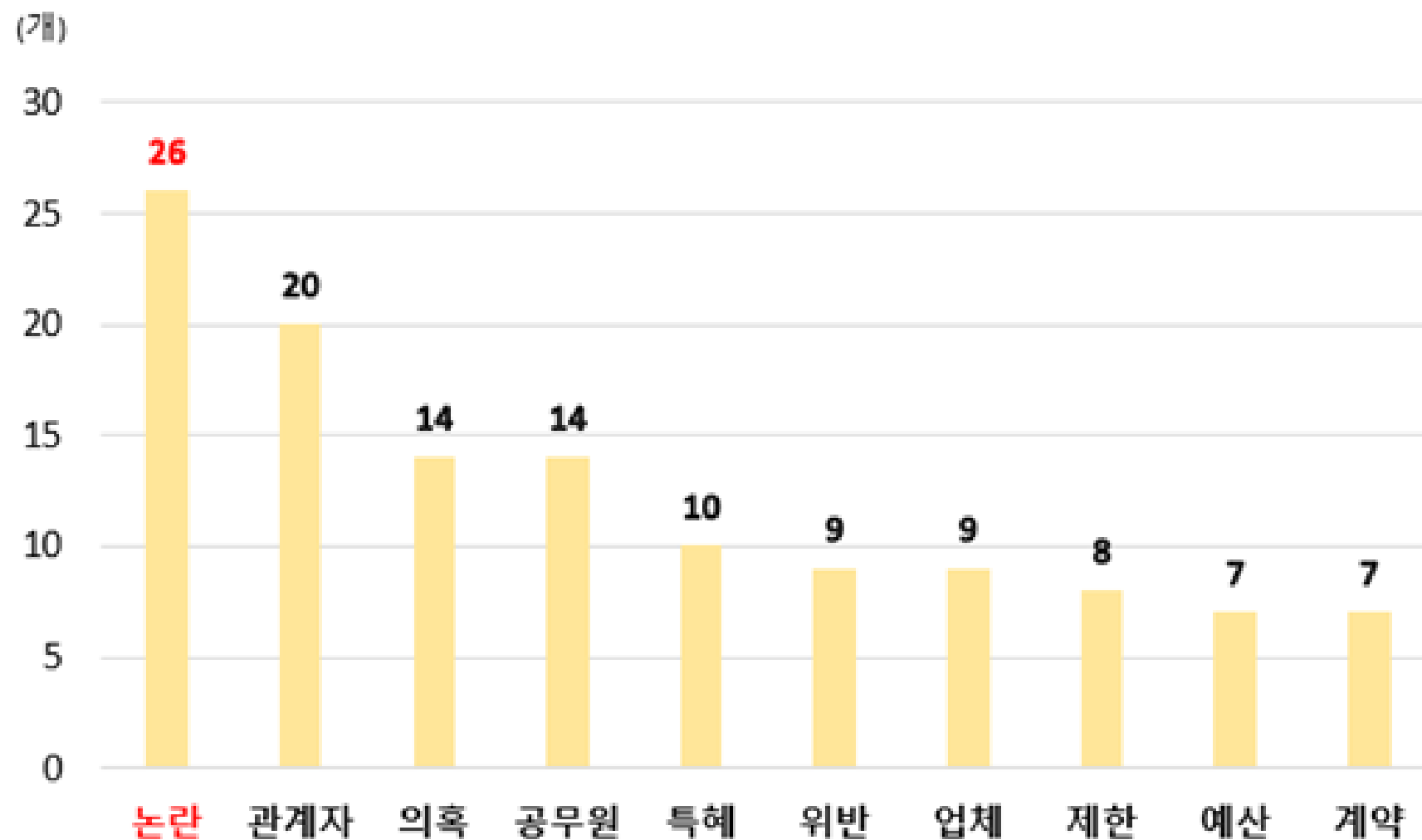
- 최근 자치사무에 대한 포괄적, 사전적 감사나 위법사항을 적발하기 위한 **선제감사 불허용**
 - ▶ 남양주시와 경기도 간의 권한쟁의에 대한 헌재 결정, '22.08.31'
- 감사주제의 선정 및 **신속한 감찰 대응**은 감사품질의 제고를 위한 매우 중요한 문제임
- 감찰 근거를 찾기 위해 업무 담당자는 **비판기사를 직접 검색하고 수집**하는 과정을 따르고 있음
 - ▶ 업무 담당자들의 업무 과중, 효율성 저하

프로젝트 목표

1. 지자체에 대한 비판기사(지방지)를 비위 유형별 키워드에 따라 **매일 검색 및 수집하는 자동화 모델 제시**
2. **Python에 대해 알지 못하는 사람도 쉽게** 모델을 사용할 수 있도록 GUI 프로그램 개발
 - ▶ GUI: "Graphical User Interface"의 약자, 사용자와 컴퓨터 간의 상호 작용을 위한 그래픽 환경을 제공하는 인터페이스

키워드 선정

검색 키워드는 비판 기사에 가장 빈번하게 등장하는 단어로, 1차적으로 검색할 단어임.



- Python의 konlpy 라이브러리를 활용하여 명사만 추출 후, **빈도 분석**을 진행함.
- 빈도 분석은 2023년 수집된 비판기사 27개의 제목과 본문 내용을 참고함.
- 27개의 비판기사 중에서 26개의 비판기사에서 등장한 **‘논란’**이 검색 키워드가 되기에 적절하다고 판단함.

키워드 선정

비판기사 예시

'GB 훼손' 구리시 승마장에 학생체험 몰아주기 논란

입력 2023-04-11 오후 4:35

구리시가 개발제한구역을 훼손한 채 승마대회를 열어 물의(경기일보 2021년 5월11일·12일·27일자 10면)를 빚었던 승마장에서 승마체험사업을 추진해 논란이다.

11일 구리시 등에 따르면 시는 올해부터 승마체험사업을 확대하면서 참가 대상을 구리 소재 초등학생과 학교 밖 청소년은 물론 사회적 배려층, 트라우마 직업군 등 장애인과 학교폭력 피해 청소년과 위기청소년 등까지 확대했다.

승마체험을 통해 사회 적응력 및 자신감 회복을 돕고 말 관련 직업탐구 기회도 제공해 향후 진로에 대한 선택의 폭을 넓혀 주자는 게 취지다.

비판 기사에는 필수적으로 **감찰 관련 단어**와 **지역명**이 포함되는 것을 확인함.

기본 키워드(163개)

감찰 관련 단어와 특정 분야의 세부 업무명

=> 특혜, 의혹, 비리, 혐의, 부당, 부정, 수의계약, 건축허가, 재난안전 등

지역키워드

17개시도 각 광역자치단체에 속한 시군구

- 경기 지역 => 수원, 성남, 의정부, 안양, 부천 등
- 인천 지역 => 중구, 동구, 미추홀구, 연수구 등

신문사 선정

복무감찰담당관에서 744개의 신문사 리스트를 제공받음

지방지: '논란' + 지역 키워드 + 기본 키워드

한 지역을 대상으로 발행하는 신문으로 17개의 시도가 포함되나
서울특별시 본청, 제주특별자치도는 자체감찰을 하므로 제외

중앙지: '논란' + 기본 키워드

전국 단위로 발행하는 신문

전문지: '논란' + 기본 키워드

특정 전문 분야를 집중적으로 다루는 신문

신문사 선별 기준

1. 일주일에 1번 이상 업데이트를 하는가?

=> 일정 기간 이상동안 기사를 발간하지 않은 신문사는
죽은 신문사로 간주하여 29개의 신문사 제외

2. 텍스트로 수집이 가능한가?

=> 영상 형태의 기사는 텍스트로 비판 기사 여부를 확인 불가.
유튜브, 페이스북으로 이루어진 16개의 신문사 제외

3. 수집이 허용되는가?

=> 신문사 자체적으로 크롤링이 불가하게 제한해둔 곳이 있음.
기사 내용을 불러올 수 없으므로 60개의 신문사 제외

744개 -> 637개

크롤러 제작 방법 선정

Selenium 라이브러리 이용

1. 평균 소요시간 2시간 이상
=> 감찰의 신속을 위한 프로젝트의 목적과 맞지 않음
2. 연결 오류 및 타임슬립 에러 빈번하게 발생
=> 유지보수가 어려움

적절하지 않은 방법이라고 판단

구글, 네이버 뉴스 활용

1. 평균 소요시간 30분 이상
=> 감찰의 신속을 위한 프로젝트의 목적과 맞지 않음
2. 검색한 키워드와 전혀 연관되지 않는 기사도 함께 수집되는 현상을 발견
=> 수집된 기사의 정확도 문제

적절하지 않은 방법이라고 판단

Beautifulsoup 라이브러리 이용

1. 평균 소요시간 10분 이내
2. 전체 신문사 리스트의 약 41%는 BeautifulSoup를 이용해서 수집 가능
3. 수집된 기사의 정확도 ↑

가장 적절하다고 판단!

기사 수집 과정

① 1차적으로 중심 키워드 '논란'으로 검색하여 모든 기사 수집

| | 지역 | 신문사 | 제목 | 날짜 | url | 본문 |
|-----|-----|--------|------------------------------------------|---------------------|--------------------------------------------------|---------------------------------------------------|
| 0 | 경기 | 용인시민신문 | "부실공사로 주택 기능 상실, 준공 승인 불허해 달라" | 입력 2023.11.09 09:40 | https://www.yongin21.co.kr/news/articleView.h... | 기흥구 보정동의 한 타운하우스가 부실시공 논란을 빚고 있는 가운데, 일부 수분양자들... |
| 1 | 경기 | 용인시민신문 | 자체 감사서 보조금 부당 지출 등 적발 | 입력 2023.11.06 11:12 | https://www.yongin21.co.kr/news/articleView.h... | 용인특례시가 최근 진행한 자체 감사에서 25건의 관리 부실·규정 위반 등을 적발해 ... |
| 2 | 경기 | 용인시민신문 | 초미 관심 용인 역삼지구 임시총회 '정상화' 험난 예고 | 입력 2023.10.17 10:30 | https://www.yongin21.co.kr/news/articleView.h... | 오는 19일 열리는 용인 역삼구역도시개발사업조합(아래 역삼조합) 임시총회를 앞두고 ... |
| 3 | 경기 | 용인시민신문 | 이 시장, 초등학교 학부모회장들과 교육환경 개선 논의 | 입력 2023.10.04 09:03 | https://www.yongin21.co.kr/news/articleView.h... | 이상일 시장은 20일 간담회를 열고 초등학교 학부모회장들과 학교별 현안 해결과 용인... |
| 4 | 경기 | 용인시민신문 | 갑질 논란 체육회 어색한 동거 | 입력 2023.09.13 09:50 | https://www.yongin21.co.kr/news/articleView.h... | 6월 용인시체육회 직원들의 오광환 회장 폭언 폭로와 함께 '회장 사퇴'를 요구하자 ... |
| ... | ... | ... | ... | ... | ... | ... |
| 776 | 인천 | 경인열린신문 | \n\n 더불어민주당 김포시 갑·을 지역위원회, "골드라인 토론... | 2023-05-05 15:11 | http://www.kiyln.kr/bbs/board.php?bo_table=to... | \n 김포시와 서울시는 무책임한 행태 사과하라!" - 더불어민주당 김포시 갑·을... |
| 777 | 인천 | 경인열린신문 | \n\n 김주영·박상혁 의원, 김포 골드라인 안전대책 마련 긴급 토... | 2023-05-03 08:29 | http://www.kiyln.kr/bbs/board.php?bo_table=to... | \n "김포 교통문제 근본 해결책은 5호선 예타 면제" 시민·전문가·정치권 한목소리... |
| 778 | 인천 | 경인열린신문 | \n\n 김포시을 시의원, 소상공·을지로위원회 일동 | 2023-04-04 18:58 | http://www.kiyln.kr/bbs/board.php?bo_table=to... | \n "경기도 꼴찌 김포페이 지원 정상화하라!" 성명서 발표- 더불어민주당 김포... |
| 779 | 인천 | 경인열린신문 | \n\n 박상혁의원, 소각장 현안 관련 주민간담회 개최 | 2023-02-09 16:39 | http://www.kiyln.kr/bbs/board.php?bo_table=to... | \n "소각장 건설, 행정의 논리 아닌 시민의 논리로 진행되어야"- 박상혁 국회의... |
| 780 | 인천 | 경인열린신문 | \n\n 인천시, 광역급행(M버스) 4개 노선 신설 확정 | 2023-10-23 16:54 | http://www.kiyln.kr/bbs/board.php?bo_table=ad... | \n 시민 편의 위해 대광위 등에 지속 요청해 4개 노선 신설 - 검단~여의도, 검... |

- 통일성 있는 코드를 구현하기 위해 데이터프레임을 이용하여 코드 구현
- 여러 개의 키워드를 동시에 검색하는 기능을 가진 신문사는 대부분 없음.
=> 통일성 있는 코드를 구현하기 위해 '논란'으로 검색하여 모든 기사 추출
- 2023년 11월 16일 기준, 경기도 지역에서 '논란'으로 검색한 결과 **781개**의 기사가 수집됨.

기사 수집 과정

② 날짜 형식 통일 후 오늘과 어제 기사만 추출

| 지 역 | 신문사 | 제목 | 날짜 | url | 본문 |
|--------|-------------|---------------------------------------------------|----------|---------------------------------------------------|---------------------------------------------------|
| 0 | 경기 수도권일보 | \n전병주 의원, 학교 전기차 충전시설 의무화...의무이행은 언제?\n입력 : 20... | 20231115 | http://www.sudokwon.com//article.php?aid=17000... | \n[유한태 기자] 서울시의회 전병주 의원(더불어민주당, 광진1)은 제321회 정례... |
| 1 | 경기 경인매일 | 경제노동위원회 김완규 위원장, "행감 이후에도 경기도의회의 지도·감독은 끝난 것이 ... | 20231115 | https://www.kmaeil.com/news/articleView.html?i... | [경인매일=최승곤기자] 경기도의회 경제노동위원회 김완규 위원장(국민의힘, 고양12)... |
| 2 | 경기 김포신문 | [발행인 칼럼] 유정복 인천시장의 망언 | 20231115 | https://www.igimpo.com/news/articleView.html?i... | 출퇴근 때마다 마주하는 지옥철과 주차장을 방불케 하는 88도로는 김포시민의 쓰라린... |
| 3 | 경기 김포신문 | [경기도 행감] 이기형 의원, "김포 서울편입 시 철도 관련 지방비 부담률 커져" | 20231115 | https://www.igimpo.com/news/articleView.html?i... | 이기형 도의원(더불어민주당, 김포4)이 14일 2023년 철도항만물류국 행정사무감사... |
| 4 | 경기 하남신문 | 하남시의회 최훈종 의원, 악성민원 피해 방지 대책 촉구 | 20231115 | http://www.ehanam.net/news/articleView.html?id... | 하남시의회 최훈종 의원(더불어민주당·나 선거구)이 11월 14일 열린 제325회 임... |
| ... | ... | ... | ... | ... | ... |
| 30 | 인천 중부일보 | '김포 서울편입' 속도내는 與... '경기북부특별자치도' 어쩌나 | 20231116 | http://www.joongboo.com/news/articleView.html... | \n로컬과 감성의 만남\n\n우리동네 집값·땅값은\n\n뉴스를 본다\n\n오직... |
| 31 | 인천 중부일보 | 尹대통령, 헌법재판관 후보에 정형식 대전고법원장 지명 | 20231116 | http://www.joongboo.com/news/articleView.html... | \n로컬과 감성의 만남\n\n우리동네 집값·땅값은\n\n뉴스를 본다\n\n오직... |
| 32 | 인천 일간경기 | 임이자·김영진 "환경 오염 피해구제, 윈스톱 서비스 통합해야" | 20231116 | http://www.1gan.co.kr/news/articleView.html?id... | [일간경기=홍정윤 기자] 임이자·김영진 국회의원은 환경 오염 피해를 입은 국민을 신... |
| 33 | 인천 일간경기 | 홍익표,尹정부 김명수 합참의장 후보자 지명 총체적 인사실패 | 20231116 | http://www.1gan.co.kr/news/articleView.html?id... | [일간경기=조태근 기자] 홍익표 민주당 원내대표는 윤석열 정부가 김명수 합참의장 후... |
| 34 | 인천 일간경기 | 민주당 "합참의장 후보자 주식·골프·학폭, 청문회 진행 필요 못 느껴" | 20231116 | http://www.1gan.co.kr/news/articleView.html?id... | [일간경기=홍정윤 기자] 김명수 합참의장 후보자의 인사 검증 청문회가 파행됐다. 국... |

- 각 신문사마다 사용하는 날짜의 형식이 다르므로 'YYYYMMDD'와 같은 형태로 통일
- 매일 지자체 비판기사를 확인한다는 해당과의 요청에 따라, 오늘과 어제 올라온 기사만 추출
- 2023년 11월 16일 기준으로 필터링한 결과, 기사는 781개 -> 35개로 나타남

기사 수집 과정

③ 본문 내용에 기본 키워드, 지역 키워드를 포함하는 기사만 추출

- 경기 지역 지방지의 검색키워드는 **'논란' + 지역 키워드 + 기본 키워드**
=> 본문에서 지역 키워드와 기본 키워드를 포함하는 기사만 추출
- 필터링한 결과 기사는 35개 -> **20개로** 나타남
- 해당 기사는 법카 지침 위반 논란으로 감찰과 관련성이 있음을 확인

[경상원 행감] 도시장상권진흥원장 법카 지침 위반 논란

▲ 김인창 기자 | © 송인 2023.11.15 09:34

| "집 근처서 전체 사용금액의 27% 사용"

[일간경기=김인창 기자] 경기도 산하 공공기관장이 업무 추진비 법인카드를 지침을 위반해 부적절하게 사용했다는 지적이 나왔다. 더욱이 맥주집이나 치킨집 등 집 근처에서 전체 사용금액의 27%를 사용했으며 일부는 개인별 한도 금액을 넘겨 질타를 받았다.



경기도의회 경제노동위원회에서 14일 열린 경기도시장상권진흥원에 대한 행정사무감사에서 의원들의 질의에 조신원장이 답변을하고 있다.(사진=김인창 기자)

이같은 내용은 경기도의회 경제노동위원회 이용욱(민주당, 파주3)의원이 11월14일 열린 경기도시장상권진흥원(경상원)에 대한 행정사무감사 질의에서 드러났다.

이용욱 의원은 이날 조신 경상원 원장을 상대로 취임 이후 법인카드 사용의 적절성을 따져 묻고 업무 추진비 27.47%를 성남 집 근처에서 사용했다며 자체 감사를 통해 적절성을 파악해 보고하도록 요구했다.

이어 이 의원은 “지난해 12월 취임해 올 8월까지 사용한 내역을 보면 총 2550만원을 썼는데, 이 중 27.47%에 해당하는 67건에 700만원 가량을 성남 자택 근처에서 사용했다”고 지적했다.

또한 “업무 추진비 사용 규정 제3장 제27조에 맥주 전문점이나 생맥주집, 선술집 이런 데서 쓰지 않게 돼있는데도 (맥주집)치어스라든지 무슨 일반적인 치킨집 이런 데서 사용하것 아니냐”고 몰아세웠다.

이와관련 조신 원장은 “모르고 썼다가 반납했다. 지적에 유념하겠다”고 답변했다.

이 의원은 계속해서 이어진 질의를 통해 “습관적으로 집 근처에서 너무 많이 사용했다. 내용을 좀 분석해 보면 이른바 흔히 얘기하는 1차 드시고 2차 가서서 또 드시고 그래요. 같은 날 시차를 두고 사용한 금액을 인원수로 계산을 해 보면, 5만2000원, 4만원, 4만9000원, 3만8000원 이래요, 이거는 좀 규정을 위반한 부분이 있다고 볼 수밖에 없다”고 재차 추궁했다.

그러면서 “7시 반에 집 근처 라면집에서 2만7000원 이렇게 쓰신게 있는데 회의를 목적으로 만나서 업무 추진비로 업무 관련성이 있다고 보기에는 참 애매하다”며 감사실장에게 내부 검토를 요구했다.

이에 조 원장은 “(경상원이) 양평에 있다 보니 주로 성남에서 미팅이 이루어진 경향이 있었으며, 여건상 불가피한 부분이 있다”고 해명했다.

Program

GUI 프로그램

Python에 대한 이해가 없는 이용자들도 GUI 프로그램을 통해 쉽게 모델을 이용 가능!

① 경기

기사 수집이 완료되었습니다.

④ Reset

| | 신문사 | 제목 | 날짜 | url |
|---|------|----------------|----------|----------------|
| 1 | 경인매일 | 경제노동위원회김완... | 20231115 | ③ https://... |
| 2 | 김포신문 | [발행인칼럼]유정복인... | 20231115 | https://... |
| 3 | 김포신문 | [경기도행감]이기형의... | 20231115 | https://... |
| 4 | 하남신문 | 하남시의회최훈종의... | 20231115 | http://... |
| 5 | 하남신문 | 박선미하남시의원,하... | 20231115 | http://... |
| 6 | 중부일보 | 문체위에결소위,지역... | 20231115 | http://... |
| 7 | 중부일보 | '메가시티서울'두고엿... | 20231115 | http://... |
| 8 | 기호일보 | 발전소와행정구역 | 20231115 | http://... |
| 9 | 일간경기 | [경상원행감]도시장상... | 20231115 | http://www.... |

① **지역별, 신문사별 버튼**: 버튼을 클릭하면 버튼에 해당하는 지역 또는 신문사의 비판 기사를 불러옴

② **데이터 프레임**: 수집이 완료된 비판 기사는 데이터 프레임 형식으로 나타남.

③ **하이퍼링크**: 데이터 프레임을 통해 수집된 기사의 제목을 먼저 확인이 가능하고 이후 본문 내용을 확인하고 싶을 때, url을 클릭하면 해당 기사로 연결됨.

④ **리셋 버튼**: 데이터 프레임을 초기화할 수 있음.

Demonstration

프로그램 시연



① 원하는 지역 또는 신문 버튼 클릭하기



② 신문 수집이 완료되면,
아래에 데이터 프레임 형태로 기사가 나타남



③ 크기 조절을 통해 기사의 제목 확인 후,
원하는 기사의 url 클릭하기

Demonstration

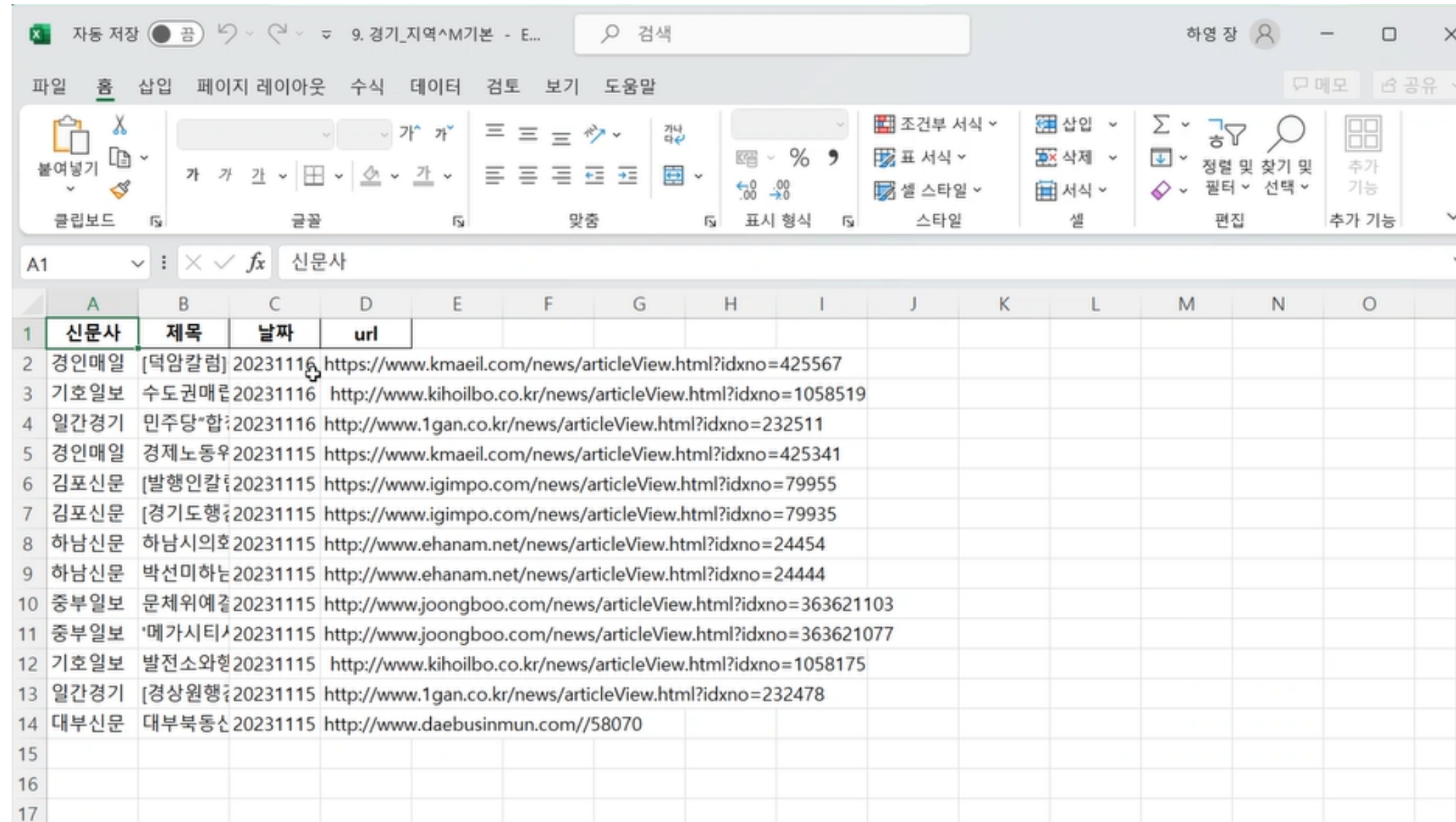
프로그램 시연



④ url을 클릭하면 해당 기사로 바로 이동 후 본문 내용 확인 가능

Demonstration

프로그램 시연



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|------|---------|----------|---------------------------------------------------------------|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 신문사 | 제목 | 날짜 | url | | | | | | | | | | | |
| 2 | 경인매일 | [덕암칼럼] | 20231116 | https://www.kmaeil.com/news/articleView.html?idxno=425567 | | | | | | | | | | | |
| 3 | 기호일보 | 수도권매 | 20231116 | http://www.kihoilbo.co.kr/news/articleView.html?idxno=1058519 | | | | | | | | | | | |
| 4 | 일간경기 | 민주당 | 20231116 | http://www.1gan.co.kr/news/articleView.html?idxno=232511 | | | | | | | | | | | |
| 5 | 경인매일 | 경제노동 | 20231115 | https://www.kmaeil.com/news/articleView.html?idxno=425341 | | | | | | | | | | | |
| 6 | 김포신문 | [발행인칼럼] | 20231115 | https://www.igimpo.com/news/articleView.html?idxno=79955 | | | | | | | | | | | |
| 7 | 김포신문 | [경기도행] | 20231115 | https://www.igimpo.com/news/articleView.html?idxno=79935 | | | | | | | | | | | |
| 8 | 하남신문 | 하남시의 | 20231115 | http://www.ehanam.net/news/articleView.html?idxno=24454 | | | | | | | | | | | |
| 9 | 하남신문 | 박선미하 | 20231115 | http://www.ehanam.net/news/articleView.html?idxno=24444 | | | | | | | | | | | |
| 10 | 중부일보 | 문체위예 | 20231115 | http://www.joongboo.com/news/articleView.html?idxno=363621103 | | | | | | | | | | | |
| 11 | 중부일보 | '메가시 | 20231115 | http://www.joongboo.com/news/articleView.html?idxno=363621077 | | | | | | | | | | | |
| 12 | 기호일보 | 발전소와 | 20231115 | http://www.kihoilbo.co.kr/news/articleView.html?idxno=1058175 | | | | | | | | | | | |
| 13 | 일간경기 | [경상원 | 20231115 | http://www.1gan.co.kr/news/articleView.html?idxno=232478 | | | | | | | | | | | |
| 14 | 대부신문 | 대부북동 | 20231115 | http://www.daebusinmun.com//58070 | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | |

⑤ 프로그램을 실행하면 자동으로 해당 날짜 파일이 만들어지고, 엑셀 파일에 기사가 저장됨.

ex) 저장경로: 기사수집 > 231116 > 지방지 > 경기.xlsx

기대효과 및 활용방안

1. 업무 효율성 향상

- 업무 담당자의 PC에 프로그램을 설치하여 **현장에서도 이용 가능**
- 매일 일일이 검색하는 과정이 줄어들어 효율적인 업무 수행 가능

2. 향후 활용도 향상

- 비판기사는 발간된 이후 삭제가 되는 경우가 빈번히 있으나 프로그램을 활용하면 날짜별로 엑셀 파일에 저장되어 다시 찾아볼 수 있음
- 비판기사 데이터가 충분히 쌓이면 텍스트 분석을 통해 **지역별 감찰 주제 발굴** 혹은 **비판기사 키워드 재정립** 등의 과정을 수행할 수 있을 것으로 기대됨

청년인턴 데이터 분석 우수사례 설명회

Thank you!

감사합니다.