

캡스톤 디자인(1) 최종 발표

Web Crawling, NLP, WordCloud를 이용한
기사 통합 검색 시스템

목차

1. 프로젝트 소개 -----	3 - 11
▪ 개발 동기, 배경, 개요	
2. 설계 및 구현 -----	12 - 19
▪ Software architecture, Dataset, Preprocessing 결과	
3. 향후 계획 -----	20 - 22
▪ 개선 필요 사항, 향후 진행 계획	
4. 참고 문헌 -----	23 - 24

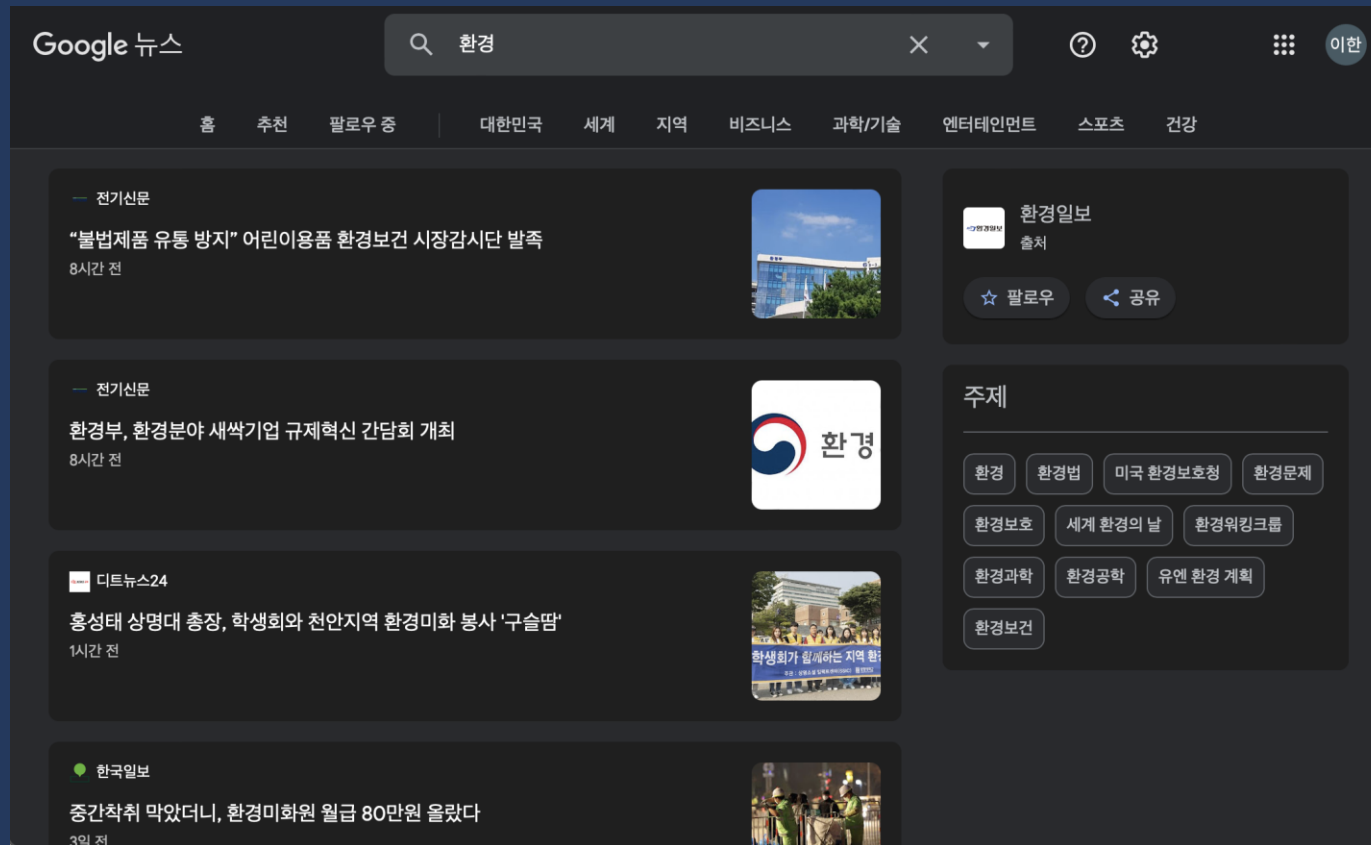


프로젝트 소개

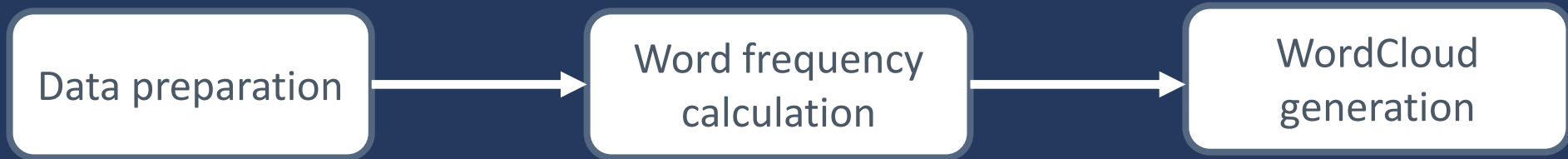
프로젝트 개발 동기



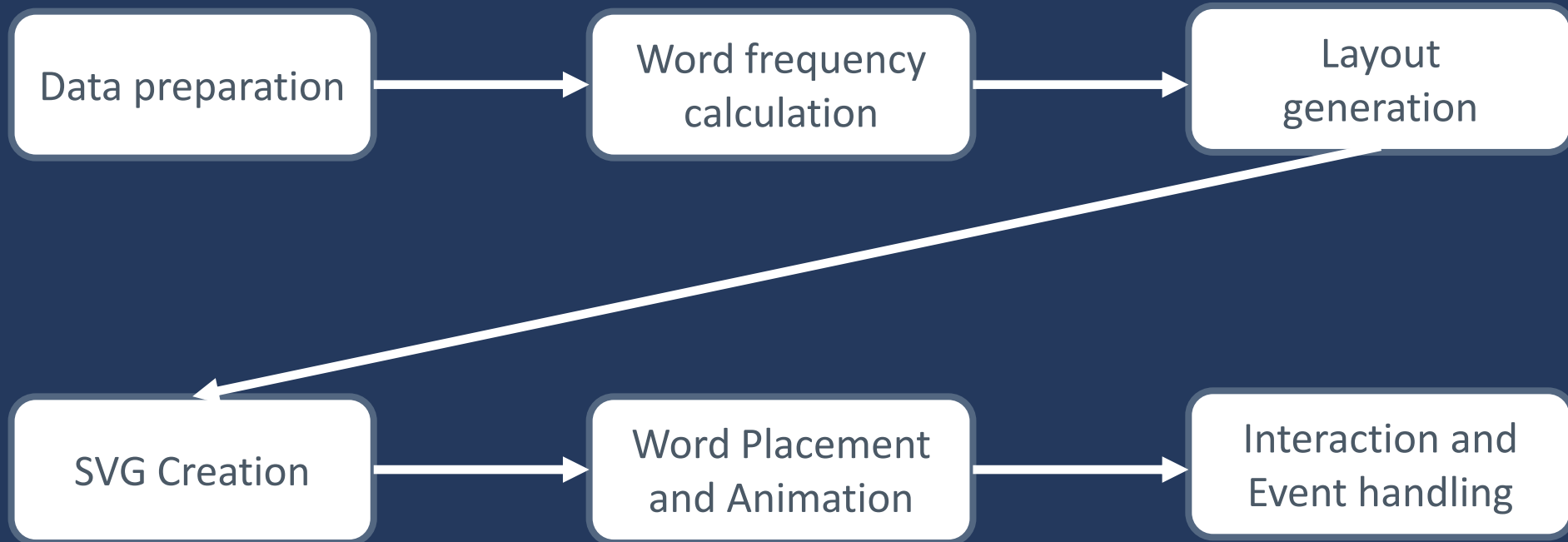
프로젝트 개발 동기



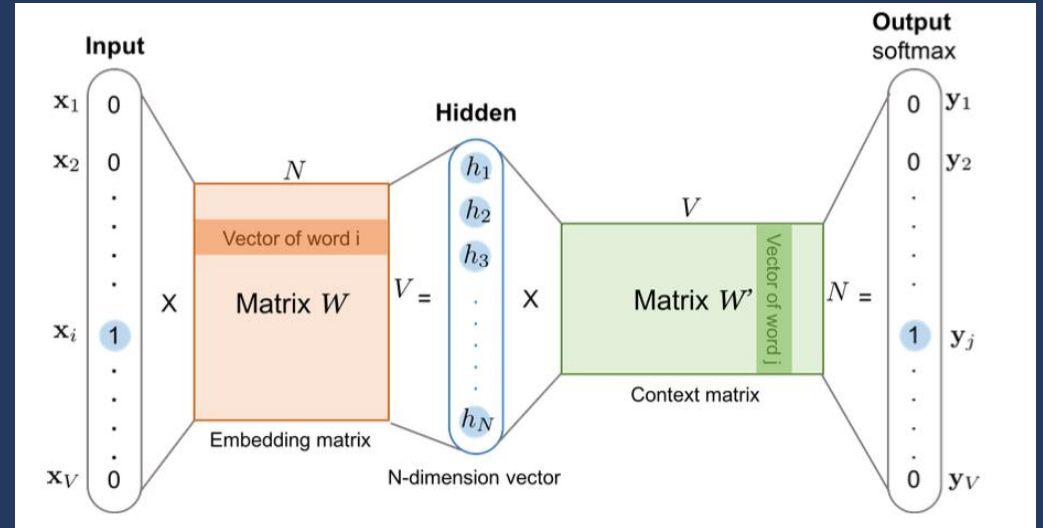
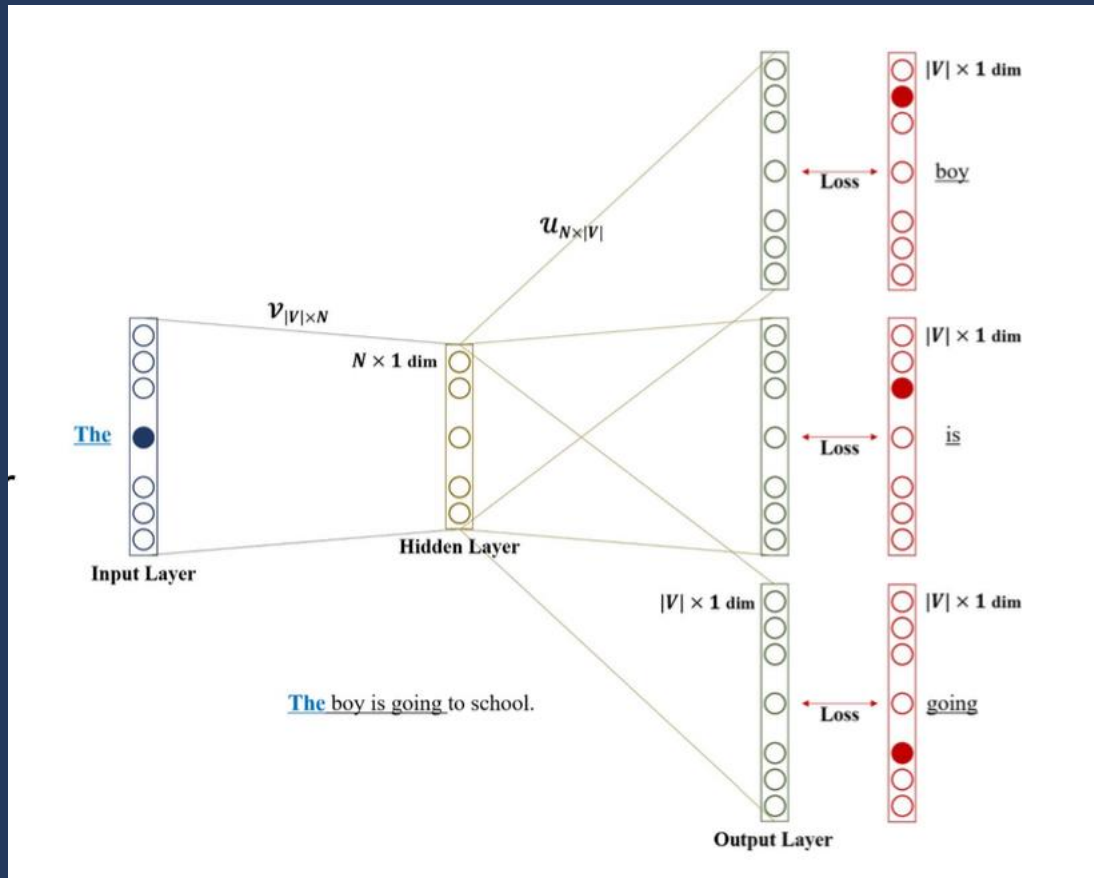
프로젝트 배경 : WordCloud



프로젝트 배경 : d3-Cloud



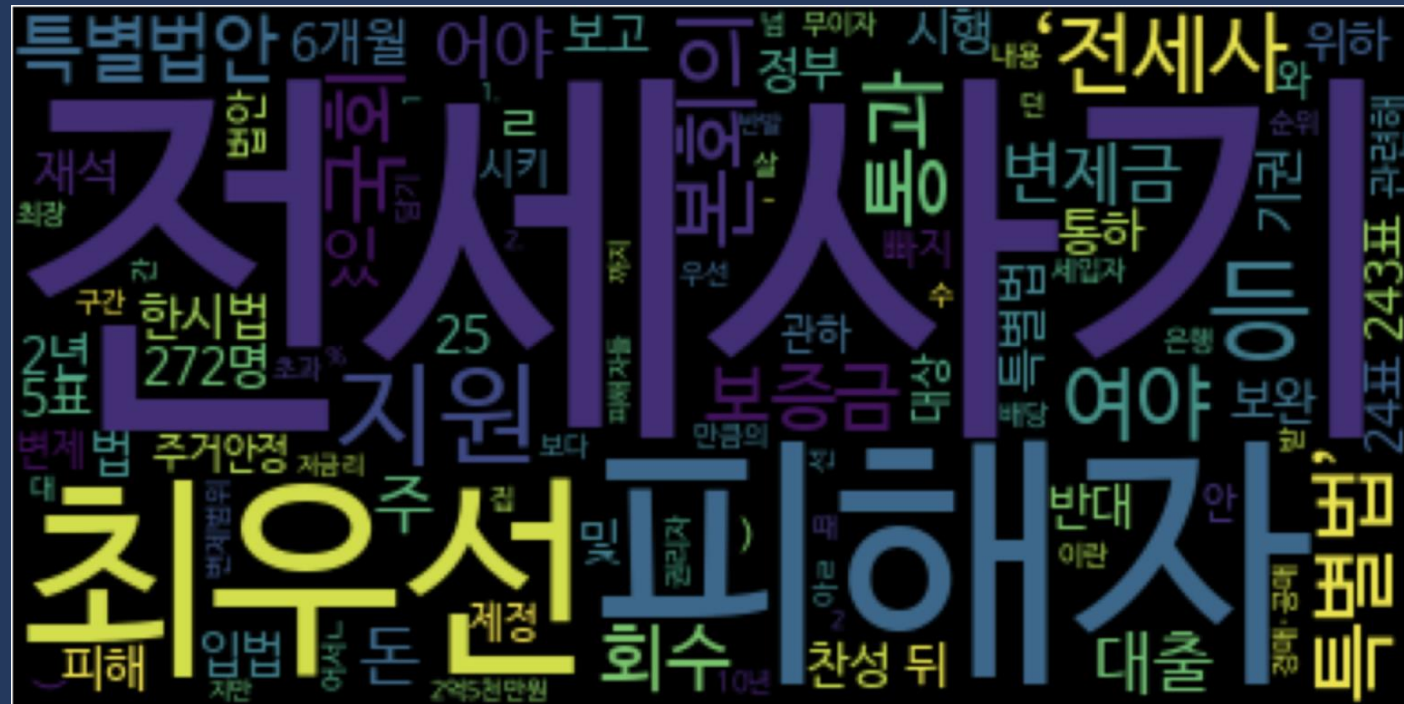
프로젝트 배경 : word2vec



프로젝트 개요



프로젝트 개요



프로젝트 개요

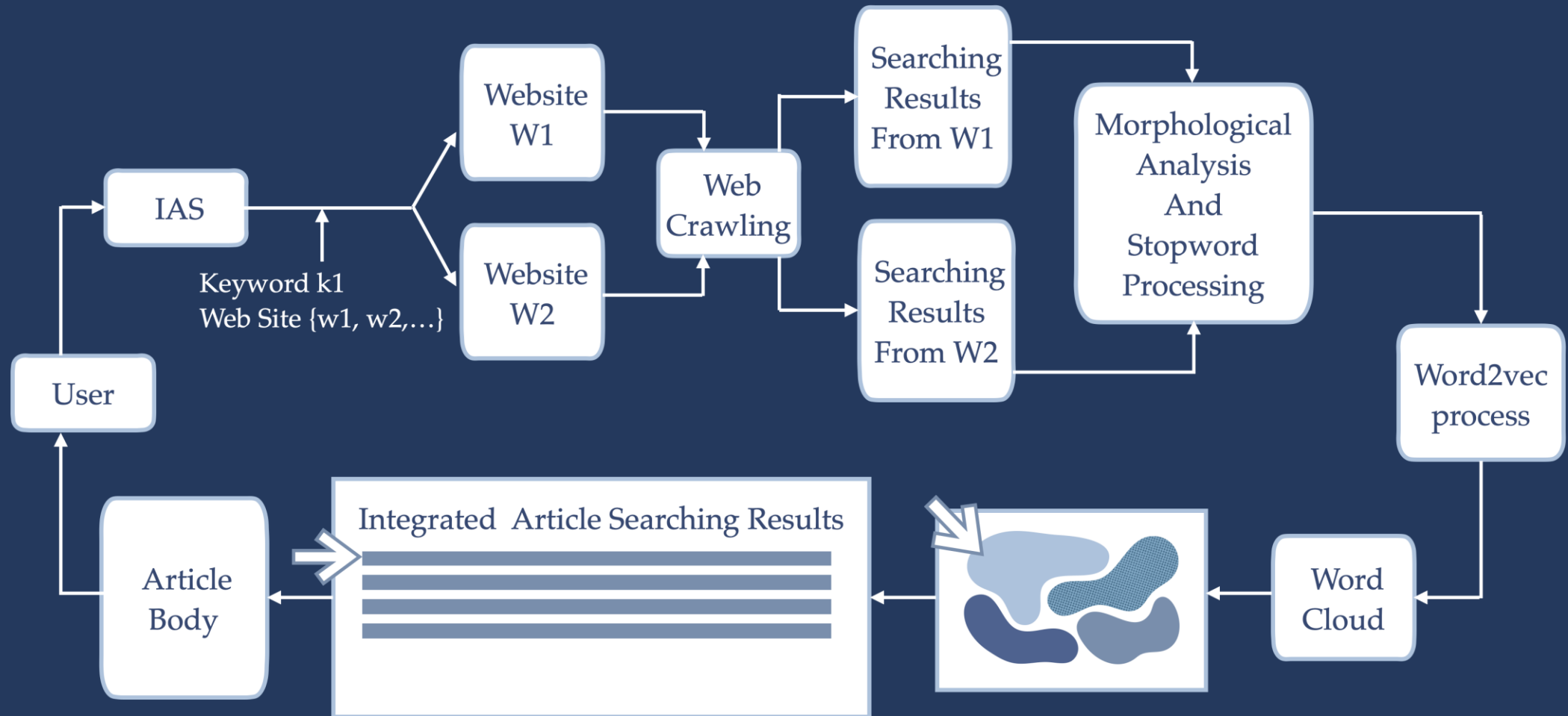
ARTICLES VIA KEYWORD



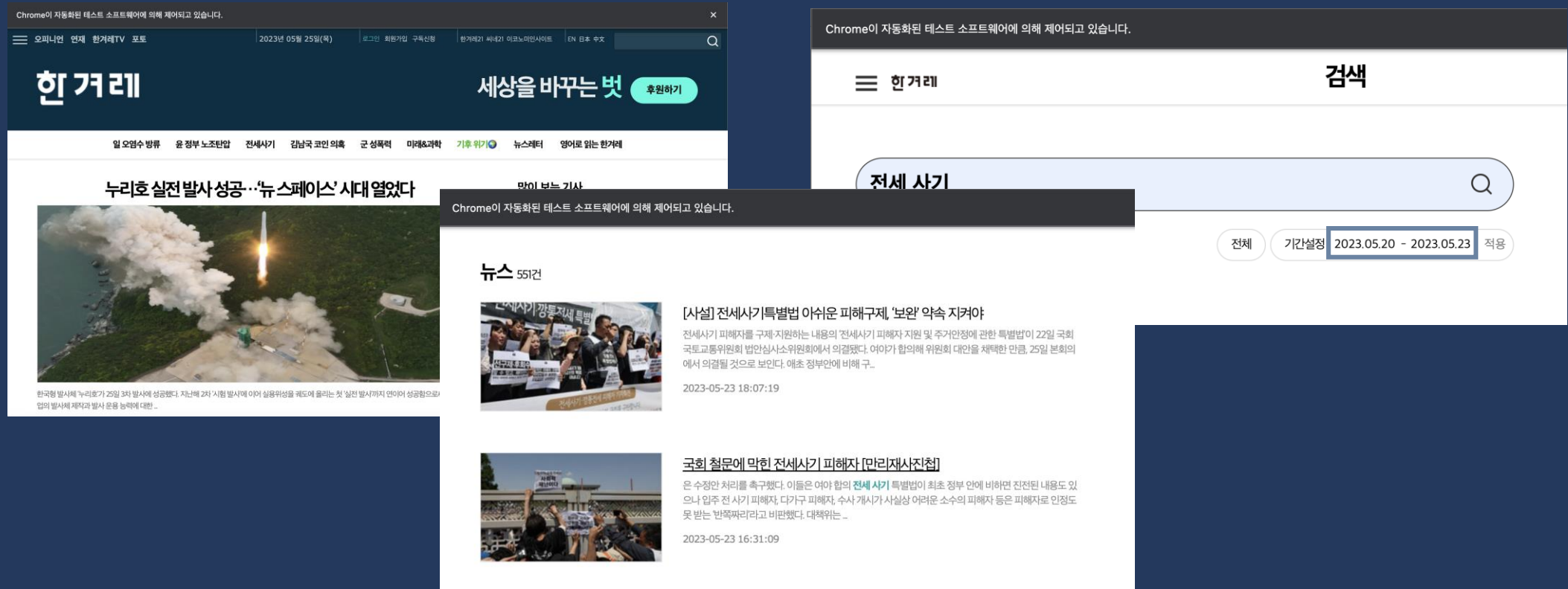


설계 및 구현

Software Architecture



Dataset: WebCrawling by Selenium



한겨레, 연합뉴스, MBC 뉴스, KBS 뉴스, 조선일보 포함 예정

Dataset: WebCrawling by Selenium

```
print(article_body)
```

['전세사기 피해를 지원하기 위한 '전세사기 특별법'이 국회 본회의를 통과했다. 특별법은 2년 한 시범으로 여야는 법 시행 뒤 6개월마다 정부 보고를 통해 보완 입법을 하기로 했다.', '국회는 25일 본회의에서 재석 272명 중 찬성 243표, 반대 5표, 기권 24표로 '전세사기 피해자 지원 및 주거안정에 관한 특별법' 제정안(특별법안)을 통과시켰다. 법안에는 보증금 회수와 관련해 최우선 변제 대상에서 빠진 전세사기 피해자에게 최우선 변제금만큼의 돈을 최장 10년간 무이자로 대출해주는 내용 등이 담겼다. 최우선 변제금이란 세입자가 살던 집이 경매·공매로 넘어갔을 때 은행 등 선순위 권리자보다 우선 배당받을 수 있는 돈이다. 여야는 최우선 변제범위를 초과하는 구간에 대해선 2억5천만원까지 1.2~2.1%의 저금리 대출을 해주기로 했다.', "하지만 피해자들은 특별법안에 반발하고 있다. 전세사기·강통전세 피해자 전국대책위원회 등은 전날 입장문을 내어 "인천 미추홀구에서 일명 '건축왕'으로 불리는 남아무개씨의 전세사기 피해자인, 5번째 전세사기 희생자가 나왔다"며 "정부는 지금이라도 선 구제 후 회수 방안을 검토해야 하며, 최우선변제금도 못받는 이들에게 최소한의 보증금 회수방안이나 주거비 지원을 해야 한다"고 말했다.', '서영지 기자']

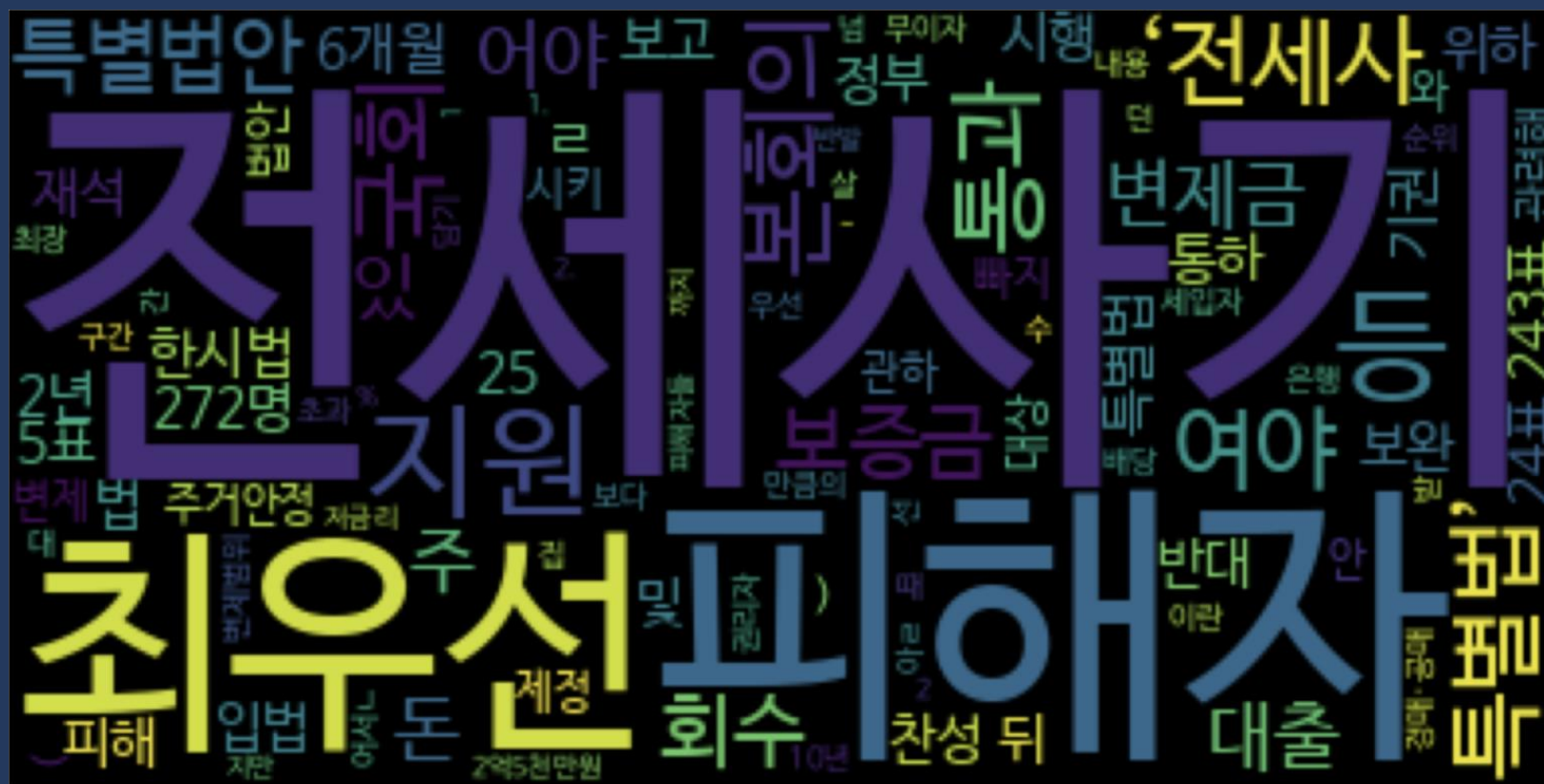
Preprocessing Result : Kkma



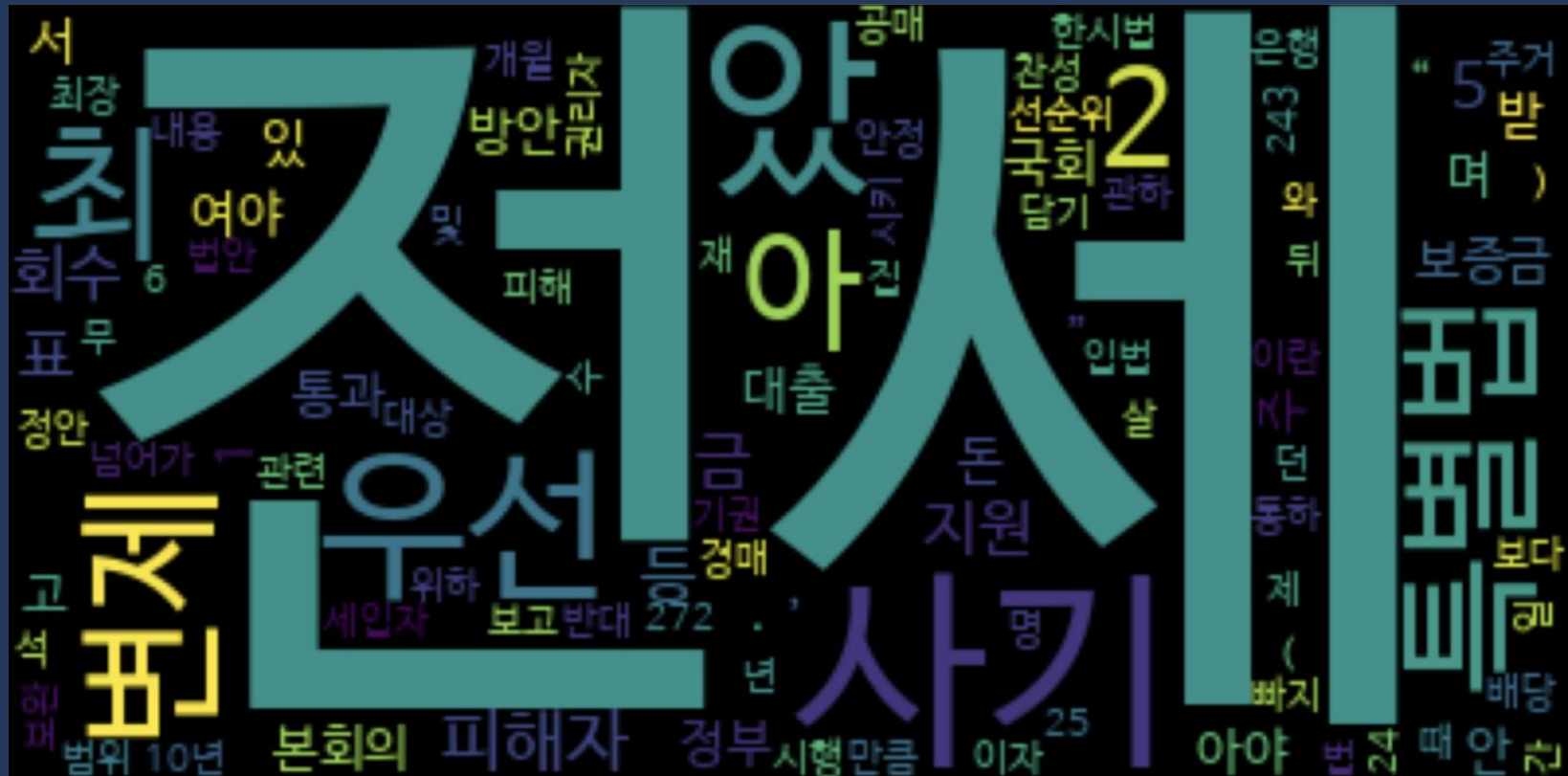
Preprocessing Result : Okt



Preprocessing Result : Hannaum



Preprocessing Result : Komoran





향후 계획

개선 필요 사항

- 뉴스 매체 분석 완료
- Preprocessing 개선
- UI 구현
- Wordcloud 생성 전에 word2vec로 유사한 data끼리 그룹화
- Clickable wordcloud python으로 구현

향후 진행 계획



참고 문헌

참고 문헌

- Clickable wordcloud built with d3-cloud
 - <https://observablehq.com/@adhitya72/clickable-word-cloud>
- 'Natural Language Processing with Transformers; Building Language Applications with Hugging Face', Lewis Tunstall, Leandro von Werra & Thomas Wolf, O'Reilly