캡스톤 디자인(2) 중간 발표

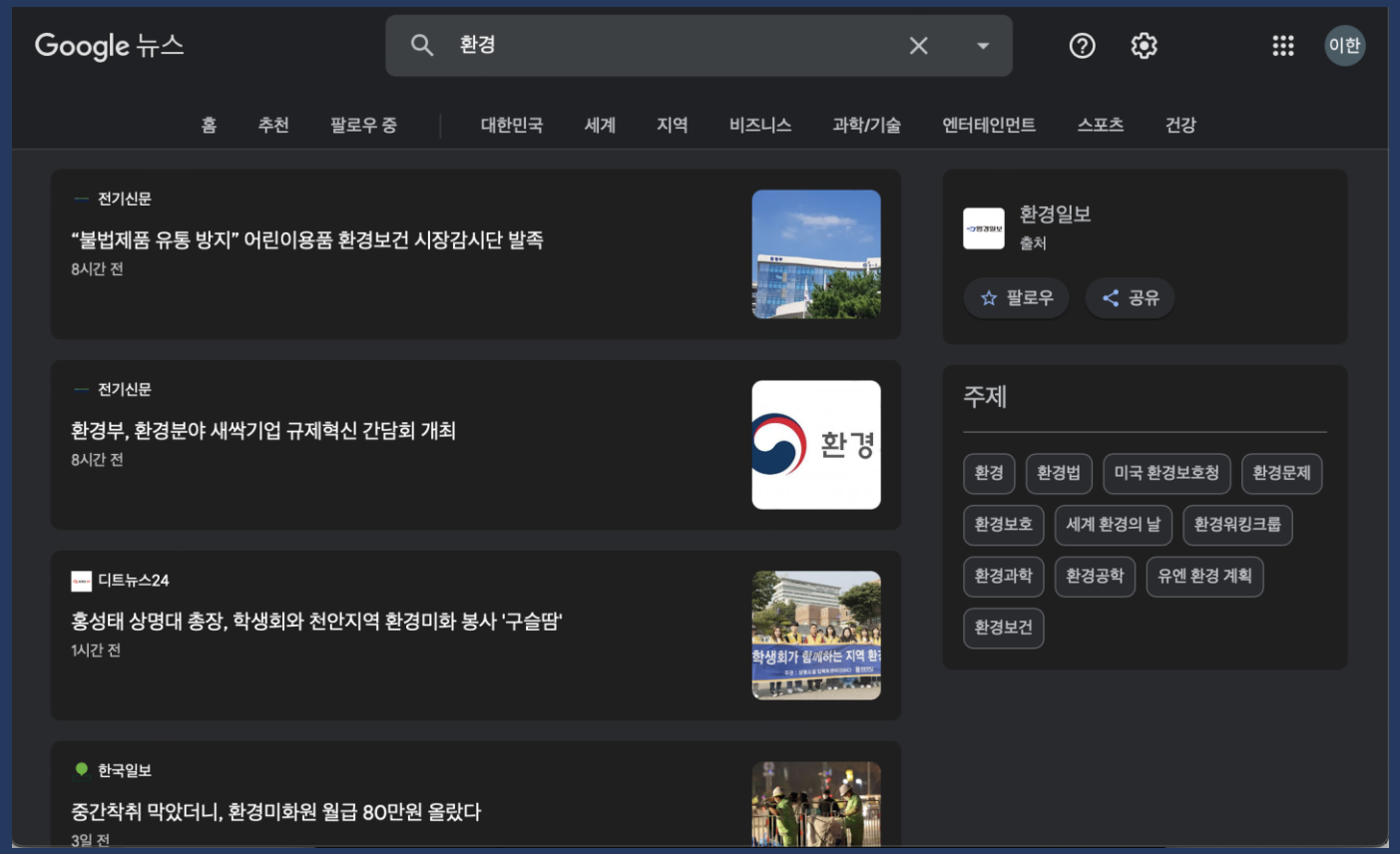
Web Crawling, NLP를 활용한
기사 통합 요약 시스템

목차

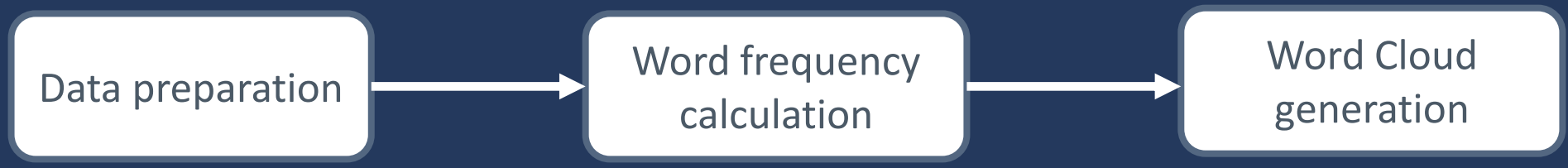
프로젝트 소개

프로젝트 개발 동기

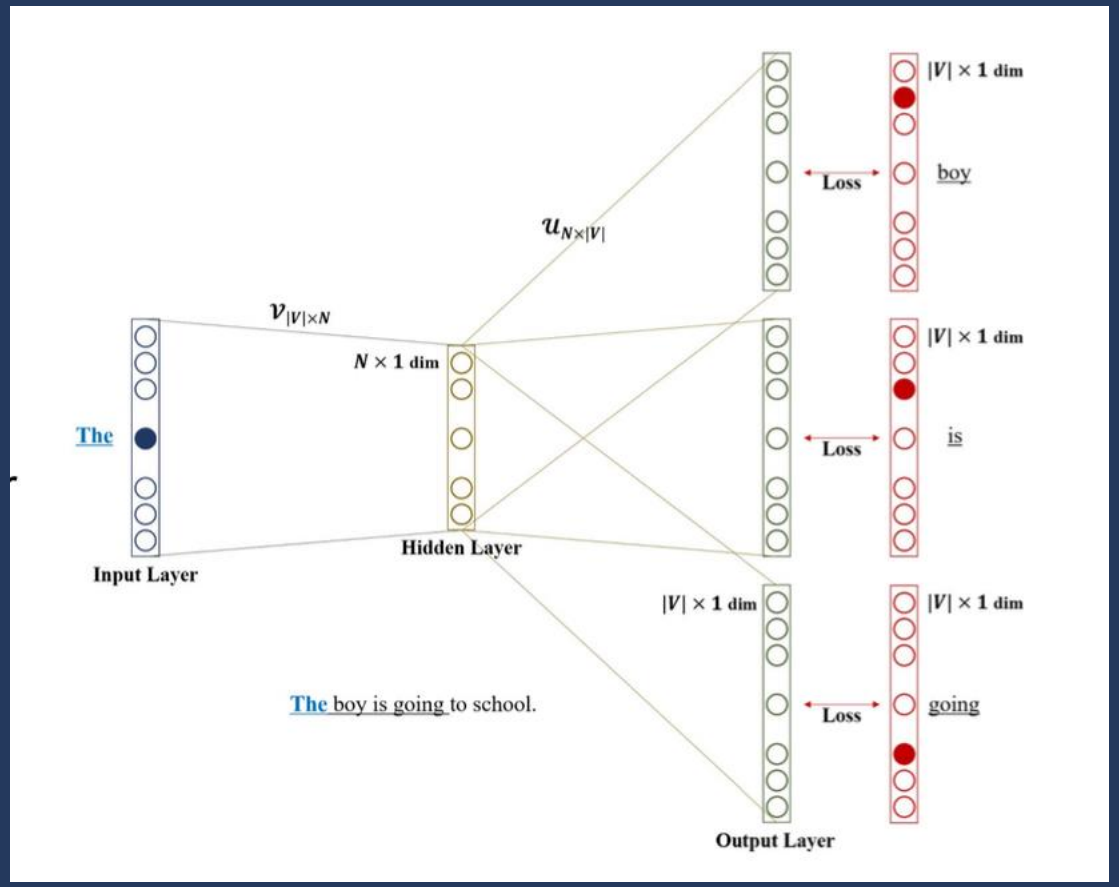
프로젝트 개발 동기

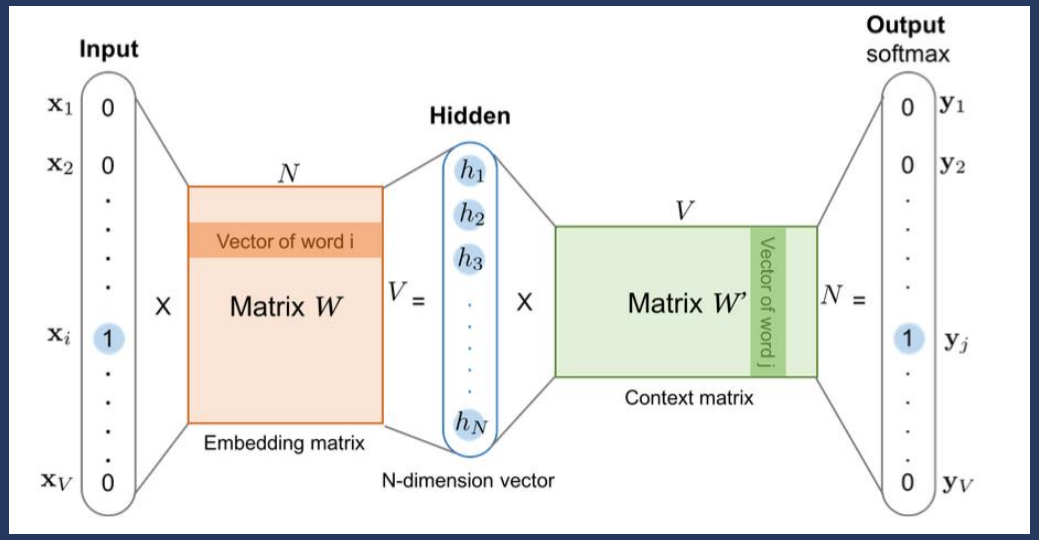

Google 뉴스
환경
이한
홈
추천
팔로우 중
대한민국
세계
지역
비즈니스
과학/기술
엔터테인먼트
스포츠
건강
전기신문
"불법제품 유통 방지" 어린이용품 환경보건 시장감시단 발족
8시간 전
전기신문
환경부, 환경분야 새싹기업 규제혁신 간담회 개최
8시간 전
환경
디트뉴스24
홍성태 상명대 총장, 학생회와 천안지역 환경미화 봉사 '구슬땀'
1시간 전
한국일보
중간착취 막았더니, 환경미화원 월급 80만원 올랐다
3일 전
환경일보
출처
팔로우
공유
주제
환경
환경법
미국 환경보호청
환경문제
환경보호
세계 환경의 날
환경워킹크룹
환경과학
환경공학
유엔 환경 계획
환경보건

프로젝트 배경 : Word Cloud



프로젝트 배경 : Word2Vec





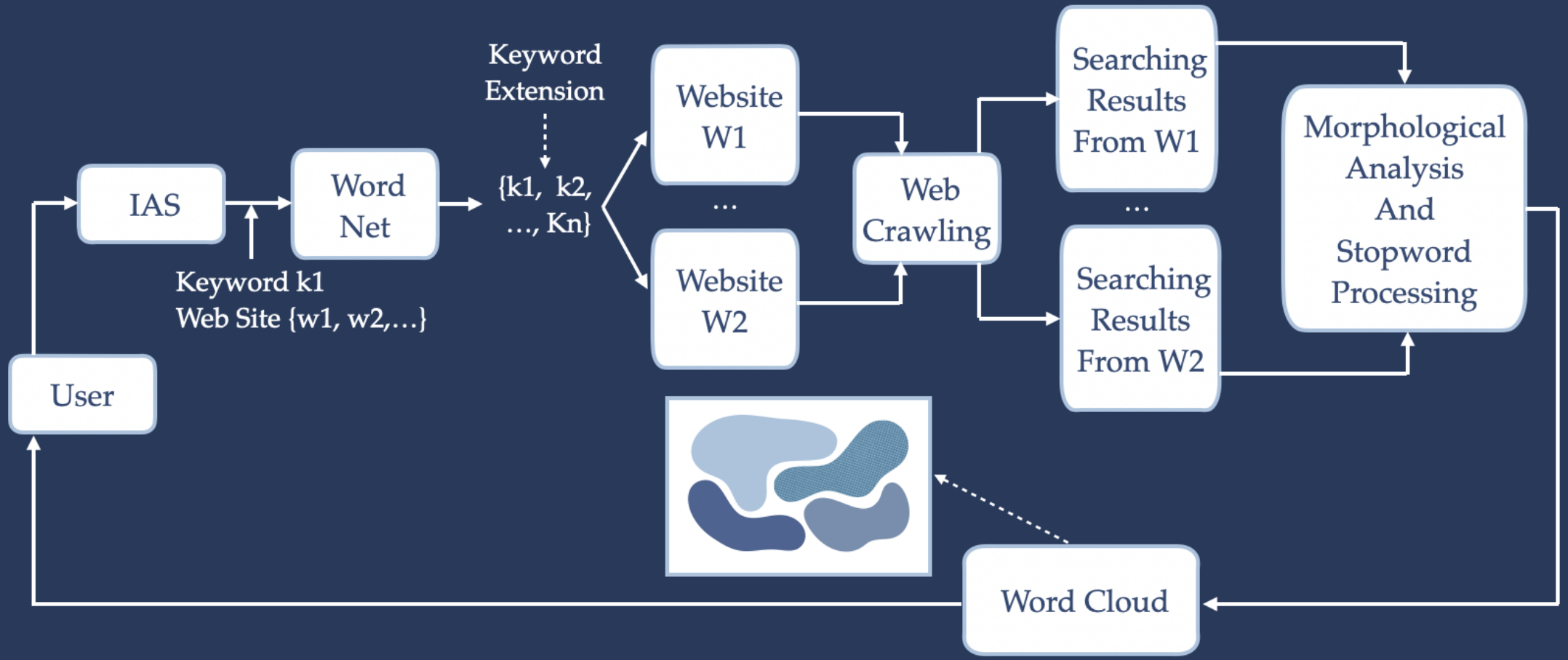
프로젝트 개요

프로젝트 개요


정우영
우즈베키스탄
황선홍
엄원상
항저우
조영욱
이강인
대회
중국
앞
높이
다시
발
전반
패스
공세
이후
주장
골키퍼
수비수
공
결승
멀티골
오르
이광연
상대
감독
골파고
침투
한국
만들
등
아다
골망
부근
골문

설계 및 구현

Software Architecture



Dataset: Web Crawling

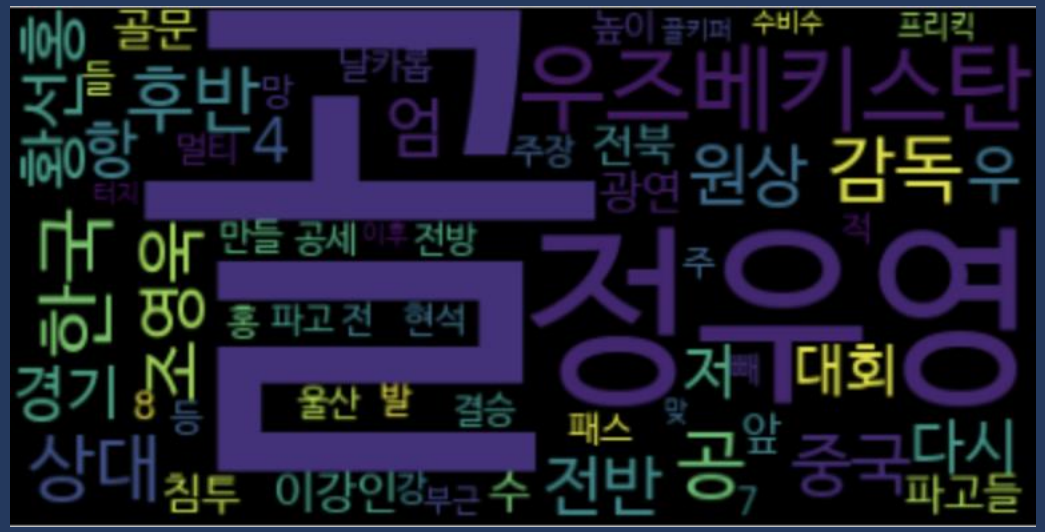
정우영의 물오른 골 감각이 또 다시 승리를 안겼다. 득점 선두 행진도 탄력을 붙였다.황선홍 감독이 이끄는 24살 이하 남자축구 대표팀은 4일 중국 항저우 황룽 스포츠센터 스타디움에서 열린 2022 항저우아시안게임 준결승전 우즈베키스탄과 경기에서 정우영의 멀티골로 2-1로 이겼다. 한국은 홍콩을 4-0으로 제압하고 결승에 오른 일본과 7일 밤 우승컵을 놓고 다툰다. 아시안게임 두 대회 연속 결승 한일전 성사다.황선홍 감독은 이날 선발 공격진에 이강인(파리 생제르맹)과 정우영(슈투트가르트), 엄원상(울산)을 배치했다. 최전방의 조영욱(서울)을 빼고는 전방 공격수들이 8강 중국전 때와는 달라졌다. 이영표 해설위원은 “최강의 멤버로 선발진을 구성했다”고 분석했다.황선홍 감독은 중원에는 중국전 때와 마찬가지로 주장 백승호(전북)와 홍현석(헨트)에게 공·수의 연결과 공 배급 등을 맡겼다. 수비진에는 설영우(울산), 박진섭(전북), 이한범(미트윌란), 황재원(대구)이 늘어섰고, 골문은 이광연(강원)이 지켰다.시작부터 속도와 패스 플레이로 날카롭게 파고든 한국의 첫골 결실은 전반 4분 나왔다. 오른쪽 측면의 엄원상이 상대 진영 배후를 파고든 뒤 낮고 강하게 공을 올렸고, 침투하던 정우영이 가볍게 발로 터치하면서 골망을 흔들었다. 정우영의 대회 6번째 골.홍현석과 엄원상이 만들어내는 침투와 이강인의 좌우 횡단 패스, 정우영의 발 빠른 침투는 반복적으로 우즈베키스탄의 골문을 위협했다. 앞선의 조영욱도 중거리포로 공세의 파고를 높였다.하지만 우즈베키스탄도 만만치 않았다. 강대강으로 맞선 우즈베키스탄은 전반 25분 벌칙구역 앞 부근에서 프리킥 반칙을 얻어냈고, 주장 자로리디노프가 때린 공은 수비수를 맞고 굴절되면서 몸을 날린 이광연 골키퍼의 손을 맞고 들어갔다. 경기장을 가득 메운 중국 관중들은 수세에 몰렸던 우즈베키스탄을 응원했고, 골이 터지자 환호했다.이후 치고받는 팽팽한 경기는 정우영의 멀티골로 다시 한국으로 기울었다. 정우영은 전반 38분 상대 골지역 왼쪽을 파고들며 상대 수비수가 미처 걷어내지 못한 공을 주워 먹듯 골망 안으로 넣으며 흐름을 가져왔다. 정우영의 대회 7호골.후반 초반은 우즈베키스탄의 공세 파고가 높았다. 한국은 후반 8분께 아크 부근에서 프리킥을 내줬고, 상대 키커의 날카로운 슈팅을 이광연 골키퍼가 정면에서 막아내면서 가슴을 쓸어내렸다.황 감독은 후반 15분께 송민규(전북)와 정호연(광주)을 투입하고 이강인과 정우영을 빼주면서 다시 팀 동력을 높였다. 상대의 거친 플레이도 이어져, 후반 17분께는 엄원상이 돌파하다 위험하게 넘어지기도 했다. 결국 전반 경고를 받았던 우즈베키스탄의 부리프가 후반 28분 조영욱을 막다가 경고누적으로 퇴장당하면서 한국은 수적 우위를 누리게 됐다.이후 한국의 조영욱과 안재준(부천) 등이 개인 능력을 발휘해 골문 앞에서 좋은 기회를 만들었지만 추가골은 터지지 않았다. 오히려 10명이 싸운 우즈베키스탄이 투혼의 경기를 펼치면서 황선홍 감독은 종료 휘슬이 울릴 때까지 안심할 수 없었다.항저우/김창금 선임기자

Preprocessing Result : Kkma

Preprocessing Result : Okt

Preprocessing Result : Hannaum

Preprocessing Result : Komoran



지난 Preprocessing Result와 비교

향후 계획

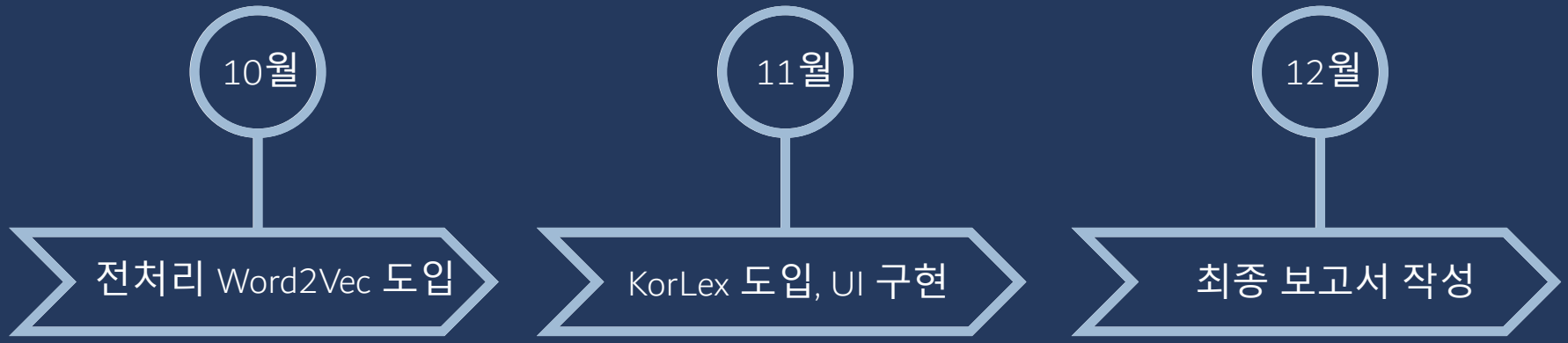
최종발표까지 예상 결과

- 전처리에 Word2Vec 도입
- KorLex 도입을 통한 검색 단어의 다양성

최종발표까지 예상 결과

- 유사성을 기반으로 그룹화한 워드 클라우드 생성 가능
- 데이터 증가 여부 선택 가능
- 증가한 데이터를 기반으로 한 워드 클라우드 생성 가능

향후 진행 계획



참고 문헌

참고 문헌

- ‘Natural Language Processing with Transformers; Building Language Applications with Hugging Face’, Lewis Tunstall, Leandro von Werra & Thomas Wolf, O’Reilly
- ‘Word Cloud Explorer: Text Analytics Based on Word Clouds’, Heimerl, Florian. 2014 47th Hawaii International Conference on System Sciences ISBN