

멘토링 자료

프로젝트 주제 선정을 위한 내용 정리 문서

조명 : (임시) 1조

조원수 : 6명

조원명 : 고은영, 박기현, 유재찬, 윤승욱, 조민우, 하정수

프로젝트 1. 뉴스 요약 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 2. 레시피 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 3. 웹소설 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 4. 운동자세 교정 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트명	바쁜 출퇴근길의 직장인들을 위한 오늘의 간추린 뉴스 AI	
프로젝트 목적	미디어의 생산 및 유통과정에 AI가 주도적으로 참여하게되어, AI솔루션(GAN)을 활용하여 변형한 콘텐츠를, 이용자가 선택한 성향에 맞춘 초개인화 서비스와 함께 제공	
기대효과	정량적 측면	<ul style="list-style-type: none">• 인건비 감소<ul style="list-style-type: none">- 사람의 간추린 뉴스의 작업 비용감소• 웹 화면 부수적인 광고수익 창출
	정성적 측면	<ul style="list-style-type: none">• 스케줄링을 통해 반복적인 탐색 행위의 불편함 해소• 개인 맞춤형 편성<ul style="list-style-type: none">- 정치, 문화 등 개인적인 성향을 반영한 언론사 , 기자선택- 관심사 위주 뉴스 섹션 선택• 정보의 공유와 확보차원에서 이용자의 능동적 구독<ul style="list-style-type: none">- 제한된 언론사 기사 뿐 아니라, 페이스북 등 정보가 제공되는 어떠한 페이지로도 리소스 확장 가능• 언론사 인력 관리 지표 제공<ul style="list-style-type: none">- 서비스의 선정적 기사, 가십성 기사의 기자 필터링, 특정 기자 구독 기능을 통해 언론사 내부 직원 KPI 지표로 활용 가능

1. 프로젝트 소개

뉴스요약 AI

예상 구현 화면

개인화 설정 페이지

언론사 구독

☒ 세계일보 ☐ 한겨레 ☐ 경향신문
☐ 중앙일보 ☐ 조선일보 ☐ 국민일보

기사 선택 기준

☐ 최신순 ☐ 조회수

기사 개수

간추린 뉴스 생성

4월 25일 화요일 간추린 아침뉴스입니다.

- 윤석열 대통령이 우리 형상으로는 12년 만에 미국을 국빈 방문했습니다. 우리 시작으로 오늘 새벽 워싱턴DC에 도착한 윤 대통령은 동료 간담회를 시작으로 내일 한미정상회담, 모레 미 의회 연설 등 20개가 넘는 공식 일정을 소화합니다.
- 백악관은 내일 미대통령상에서 최정호씨 경례를 위한 추가 방문을 논의할 예정이라고 밝혔습니다. 미 상원은 한미 동맹이 강화의 핵심 요소라며 윤 대통령의 국빈 방문을 환영하는 결의안을 발의했고, 하원도 곧 동향할 예정입니다.
- 윤석열 대통령이 방미할 때 한국과 한미관계의 안착을 위한 관제에 대해 100년 전 일로 무조건 무릎을 꿇어야 한다는 반식은 받아들일 수 없다는 입장을 밝혔습니다. 여당은 일본의 식민 지배에 반의부를 주는 것이라고 비판했습니다.
- 우리 정부가 일본을 수출 우대 국가, 이른바 화이트리스트에 다시 포함했습니다. 지난날 오랜 한일 정상회담의 후속조치로 양국 간 무역 관계를 회복하기로 한 것에 일본은 아직 상용하는 조치를 내놓진 않았습니다.

● 한편 우리 무역 본영이 적화하고 있는 수단에서 우리 교민 28명이 군용기를 타고 사우디아라비아로 무사히 탈출했습니다. 이 가운데 20명은 대형 수송기를 타고 오늘 오후 서울공항에 도착할 예정입니다.

● 한국과 미국이 41일째 남아프리카에서는 처음으로 북한 국전사 1명을 동시 체포했습니다. 북한 조선로동당 소속인 이 남성은 일회회비를 보충해 수백만 달러의 불법 자금을 세탁하고 북한의 대량살상무기, WMD 자금 조달에 관여한 것으로 알려졌습니다.

● 이번 정상회담에서는 미국 반도체 지원법에 따른 국내 기업에 피해를 줄릴 방안이 논의될 지가 큰 관심입니다. 백악관은 이번 우려에 반도 제 지원법으로 한국 기업이 오히려 혜택을 볼 것이라고 강조했습니다.

● 중국이 미국 반도 제 판매를 중국에서 금지하면 삼성과 SK하이닉스는 중국 판매를 늘리지 않겠다고. 미국 백악관이 우리 대통령실에 이런 요청을 했다는 최신 보도가 나왔습니다. 사실이라면 우리 대기업과 경제에 큰 영향을 줄 수 있어 관심이 예상됩니다.

● 윤 대통령 정부 이후 최대 규모의 국가관 송영길 전 민주당 대표는 국민과 당원들에게 송구하리면서도 의회에 대한 아직 모르는 부분이 많다고 담백을 피했습니다. 김정은 장제 한국감사원회장과 윤석열, 이석만 민주당 의원 등 핵심 인물 9명에게 대해 송구금지 조치를 하고 보금 수사에 주의를 기울이고 있습니다.

● 최근 국제 무역에서 주목받는 한국 경제를 소개하기 위해 중국 수도 베이징에 한국 경제 전문가가 문을 열었습니다. 그런데 이제 개관식이 열렸지만 최근 냉랭해진 양국 관계를 반영하듯 중국 측 인사는 거의 참석하지 않았습니다.

● 경찰이 일선에서 여사의 명예를 훼손한 혐의로 고발된 안영우 이상호 장형에 대해 검찰에 넘겼습니다. 김보라와 방은 당시 연회에서 사건을 겪었다는 의혹, 외교부 장관 부인을 유해하게 대했다는 의혹, 모두 유죄입니다. 게 경찰 불판입니다.

● 양한 미주총회 전주시기 시베리아의 70%가량이 근지상 생활 시기로 인해 소액 밀착 보호를 위한 최우선변제를 받지 못하는 것으로 확인됐습니다. 정부 대책의 실효성 논란이 나온 가운데 원화를 국외로 유출한 피해자의 제우를 탐할할 수는 없다고 선을 그었습니다.

● 저중세의 신종이 발명 식품이 기본으로 전 세계인의 사랑을 받는 올리브유 가격이 지난해보다 60% 이상 치솟으며 시장 최고가를 기록 중입니다. 이유는 스페인을 달린 최악의 가뭄때문입니다. 스페인은 전 세계 올리브유 생산의 약 절반을 공급하고 있기 때문입니다.

● 방의 총인 대통령과 만난 미국의 군번트 플랫폼 기업, 넷플릭스가 한국 콘텐츠 20억 달러, 우리 돈 3조 3천여억 원을 투자하겠다고 밝혔습니다. 넷플릭스가 한국에 진출한 2016년부터 지난해까지 투자한 총 금액의 2배에 달하는 액수입니다.

● 고지대에서 무려 10만 명 이상이 사망한 이태원 붕괴사고가 발생했습니다. 경찰이 30명의 계도 기간을 거쳐 본격적인 단속에 나섰는데, 언제 끝날 수 있고, 또 이어서 문제가 되는 건지 아직도 혼란스럽다는 반응이 많습니다.

● 오늘은 전국이 초고고 남해안과 제주도에서 비가 시작해 전국 대부분 지역에 확대되었습니다. 비는 늦은 오후 서쪽 지역부터 그치기 시작하고, 한낮 기온이 어제보다 3~4도만 떨어져 쌀쌀하겠습니다.

상세 기사 화면

언론사

제목

소제목

내용

기자명 이메일

구독 or 필터링

사용모델

BERT

- 2017 Google에서 공개한 Transformer Encoding방식의 인공지능 언어모델이며 오픈소스로 사용 용이
- Token Embeddings + Segment Embeddings + Position Embeddings 각 임베딩들의 토큰별로 모두 더하여 BERT의 attention 벡터 생성
- 문장 전체를 입력받고 양방향 학습이 가능한 방법론 사용 (Masked Language Model(MLM) + Next Sentence Prediction (NSP))

KoBERT / KPF-BERT

KoBert	KpfBert
SKT에서 BERT의 한국어 성능 한계를 극복하기 위해 개발한 모델 PyTorch, TensorFlow 기반의 API를 제공	한국언론진흥재단에서 제작한 BERT기반의 언론특화 모델 PyTorch 기반 딥러닝 API제공
PyTorch에 대한 추가적인 학습없이 모델을 사용할 수 있음 성능 향상에 있어 신문기사 데이터에 대한 추가적인 학습이 필요할 것으로 예상됨	모델 사용에 있어서 PyTorch에 대한 추가적인 학습이 필요함

3. 데이터 수집 및 활용

뉴스요약 AI

1	수집대상	특정시간 기준 언론사별, 섹션별 최신기사 00건 수집
	수집방법	언론사별 사이트로부터 데이터 크롤링
	참고 URL	https://news.naver.com/
2	수집대상	추가적인 학습에 필요한 언론 데이터
	수집방법	모두의 말뭉치에서 제공하는 신문말뭉치 2022, 2021
	참고 URL	https://corpus.korean.go.kr/request/reasetMain.do?lang=ko#down

프로젝트 1. 뉴스 요약 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 2. 레시피 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 3. 웹소설 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 4. 운동자세 교정 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트명	이미지분석을 통한 레시피 추천 AI	
프로젝트 목적	1인가구를 위한 식재료 활용 및 영양 관리 서비스	
기대효과	정량적 측면	<ul style="list-style-type: none">• 온라인 스토어와 연계하여 구매 유도
	정성적 측면	<ul style="list-style-type: none">• 식자재를 최대한 활용하여 음식물 쓰레기 배출량 감소• 식단 관리를 통한 국민 건강 증진

1. 프로젝트 소개

레시피추천 AI

예상 구현 화면



사과
청포도
포도
감
수박
풋사과

추가하기

가지고 있는 재료

재료 추가

+

토마토 -3일

계란 -3일

돼지고기 -3일

소고기 -3일

가지고 있는 재료로 만들 수 있는 레시피 목록

토마토 | 계란 | 돼지고기

사진

토마토 계란 돼지고기 볶음

난이도 하 | 410 Kcal

토마토 | 계란

사진

토마토 계란 볶음

난이도 하 | 410 Kcal

토마토

사진

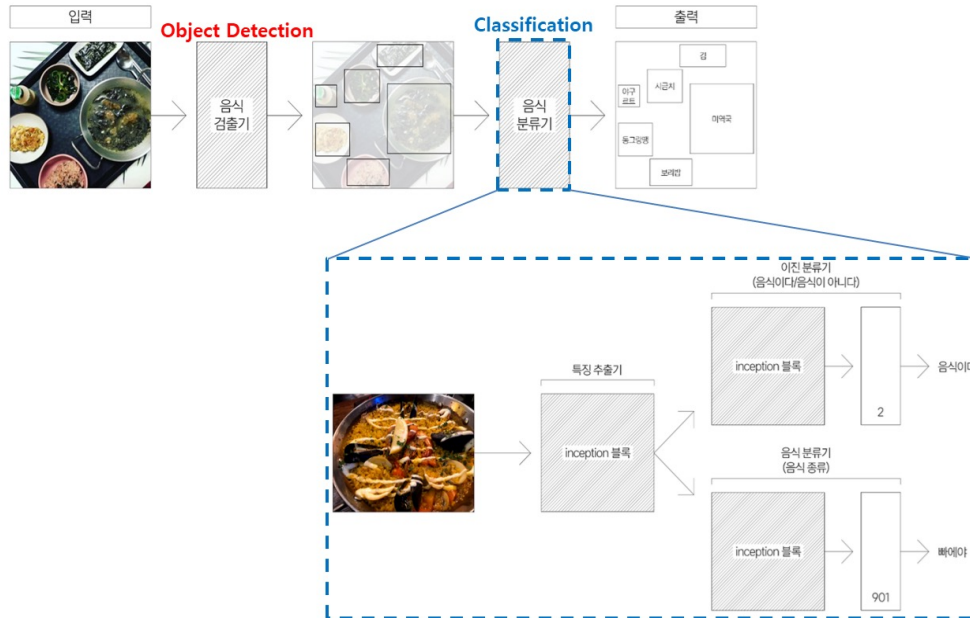
토마토

난이도 하 | 410 Kcal

2. 필요기술

레시피추천 AI

사용모델



Object Deteciton

YOLO V7

- 높은 성능을 유지하며 빠른 속도로 객체를 검출할 수 있도록 고안된 모델
- ELAN, RepConv 구조 도입
- object detection task에 맞는 scaling 방법 제안

Classification

Efficient net

- Compound Scaling을 사용해 파라미터 수를 줄이면서도 더 높은 정확성 달성
- Efficient Block 구조 도입
- EfficientNet은 B0부터 B7까지 다양한 크기의 모델 제공

Pytorch 기반 코드

3. 데이터 수집 및 활용

레시피추천 AI

1	수집대상	음식 이미지 및 영양정보 텍스트 400종 항목 (항목당 약 2천장)
	수집방법	AIHUB 제공 데이터 크롤링
	참고 URL	https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=74
2	수집대상	쇼핑몰 리뷰 이미지
	수집방법	TBD
	참고 URL	TBD
3	수집대상	음식 및 식재료의 100g당 영양정보 (5만개 데이터 제공)
	수집방법	오픈데이터 활용 (전국통합식품영양성분정보표준데이터)
	참고 URL	TBD
4	수집대상	만개의 레시피 데이터 (약 12만개 레시피 데이터 제공)
	수집방법	오픈데이터 활용
	참고 URL	https://kadx.co.kr/product/detail/0c5ec800-4fc2-11eb-8b6e-e776ccea3964

프로젝트 1. 뉴스 요약 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 2. 레시피 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 3. 웹소설 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 4. 운동자세 교정 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트명	웹소설 추천 AI	
프로젝트 목적	사용자가 본 웹소설을 기반으로 흥미를 가질 만 한 웹소설 추천	
기대효과	정량적 측면	<ul style="list-style-type: none">• 사용자가 더 오래 서비스를 이용하도록 하여 웹소설 구매량 증가• 배너 광고를 통한 수익 증대
	정성적 측면	<ul style="list-style-type: none">• TBD

1. 프로젝트 소개

웹소설추천 AI

예상 구현 화면

Contents
Based

장르

엔딩

작가

분량

~100p

101~200p

201p~

Else

재미있게 본 소설

해리포터

연금술사

셜록홈즈

신

죽음

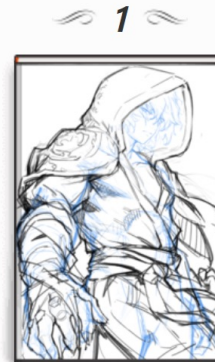
재미있게 본 소설

나미아
암화물
기적

하
일
빈!

어린 왕자

...



퓨전, 판타지
글러먹은 빙의
목마

지수 199



현대판타지
미국 피지컬 천재의 이증...
용린龍麟


지수 198



현대판타지, 판타지
시간을 달리는 소설가
피아조아

지수 198

사용알고리즘



**Content
Based**

소설 자체의 연관성 기반,
데이터 직접 수집 후 생성 가능

**Collaborative
Filtering**

사용자간의 연관성 기반,
수집된 데이터 사용 필요

**Latent
Factorization**

숨어있는 영향요소 반영 가능,
수집된 데이터 사용필요

1	수집대상	컨텐츠 기반 추천 시스템 사이트의 평점 데이터
	수집방법	Movielens dataset 활용
	참고 URL	https://www.kaggle.com/c/movielens-100k

목차

프로젝트 1. 뉴스 요약 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 2. 레시피 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 3. 웹소설 추천 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트 4. 운동자세 교정 AI

- 프로젝트 소개
- 필요기술
- 데이터 수집 및 활용

프로젝트명	운동자세 교정 AI	
프로젝트 목적	운동자세를 이미지 판독하여, AI가 자세를 교정해줌으로써 개인운동의 편의성과 자율성을 제공	
기대효과	정량적 측면	<ul style="list-style-type: none">• 헬스장 등 운동 시설 이용권 및 PT서비스 등록 비용 Saving
	정성적 측면	<ul style="list-style-type: none">• 시간 제약 제거• 운동 측정 통계량을 통한 운동관리• 잘못된 자세로 인해 유발 될 수 있는 부상의 위험도 감소

1. 프로젝트 소개

운동자세 교정 AI

예상 구현 화면

TBD



사용모델



TBD



TBD

1	수집대상	피트니스 자세 이미지
	수집방법	yoga -82 Data Set 활용
	참고 URL	https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=real&dataSetSn=231

E.O.D