



SUMEET



AI summary meeting assistant

MUD Team

201601364 박주영(T)

201600599 김아연

201503665 홍승환

201701138 김혜원

목차

01 서비스 개요

02 개발 배경 및 필요성

03 소개 및 가치 제안

04 시장규모 분석

05 제품 수익 모델

06 향후 관리 및 계획

07 업무 분담

MUD Team

Meeting Using DeepLearning

인공지능 기반 **자동 요약** 서비스

SUMeet은 인공지능을 통하여
자동으로 텍스트를 요약하는 서비스 입니다.

SUMEET

AI summary meeting assistant



2. 개발 배경 및 필요성

개발 배경 및 필요성



주요 요약콘텐츠 서비스

분야	사업자	서비스명	주요내용	
시사	뉴닉	뉴닉	주요시사이슈정리	자료:각사
책	밀리의 서재	밀리의 서재	베스트셀러를 음성으로 30분 요약해 제공	
TV방송	알려줌	알지	TV시사·교양·보도 프로그램을 5분으로 정리해 제공	
영화	1인 미디어	각종 유튜브 채널	인기영화 5~20분으로 축약한 동영상 제공	
스포츠	네이버·엔씨소프트	야구경기 하이라이트	야구경기 주요 장면을 AI로 정리해 제공	

모바일 앱 평균 이용시간 비교

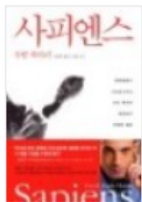
2018년 Vs. 2019년

모바일 앱	2019년 1월	2018년 1월	성장률
 유튜브	1:26:02	1:20:27	▲6.9%
 리디북스	1:12:59	0:41:09	▲77.4%
 카카오페이지	0:48:44	0:40:23	▲20.7%
 밀리의 서재	0:52:12	0:29:17	▲76.8%
 인스타그램	0:24:57	0:23:36	▲5.7%

(자료) 랭키닷컴, (이미지) 구글 플레이 스토어

지난 6월 채용 정보 제공업체 잡코리아와 알바몬이 직장인과 아르바이트 직원 1628명을 설문조사한 결과 직장인의 64.5%와 아르바이트 직원의 60.9%가 '타임푸어족(시간 빈곤층)'이라고 대답했다. 정보기술(IT)업계 관계자는 "휴대하는 스마트폰의 성능이 향상되고 무선 데이터 비용이 저렴해지면서 각종 요약 콘텐츠에 대한 수요도 커졌다"고 말했다.

□ □ □ AI □ □ □ □
□ □ □ □ □ □ □ □ ?



사피엔스 (유인원에서 사이보그까지, 인간 역사의 대담하고 위대한 질문)

지식인의 서재 추천

유발 하라리(작가) 저 | 조현욱(신문인) 역 | 김영사 | 2015.11.24.

★★★★★ 8.91 | 네티즌리뷰 [1,552건] | 도서구매 22,000원 → 19,800원 (-10%)

eBook구매 12,960원

소개 이제 우리는 무엇을 인간이라고 할 것인가지금으로부터 10만 년 전, 지구에는 호모 사피엔스뿐만 아니라 네안데르탈인, 호모 에렉투스 등 최소 6종의 인간 종이 살아 있었다. 이후 호모 사피엔스 종만이 유일한 승자로 지구상에 살아남게 되었고, 이제 그들은 신의 영역까지 넓히고 있다. 『사피엔스』는 이처럼...



Sam Van Aken
How one tree grows 40 different kinds of fruit

11:24



3. 서비스 소개 및 가치 제안

SUMEET Services

SUMeet기술을 통해 제공받을 수 있는 서비스입니다.



한글 전처리

한글 문서를 자연어 처리 및 머신 러닝을 이용하기 위해 처리



문서 소재 분류

여러 주제의 문서가 함께 있을 때, 문단을 나누는 기준으로 소재를 분류 시킨다.



키워드 추출

문서 속 핵심 단어를 한 눈에 보여줌.

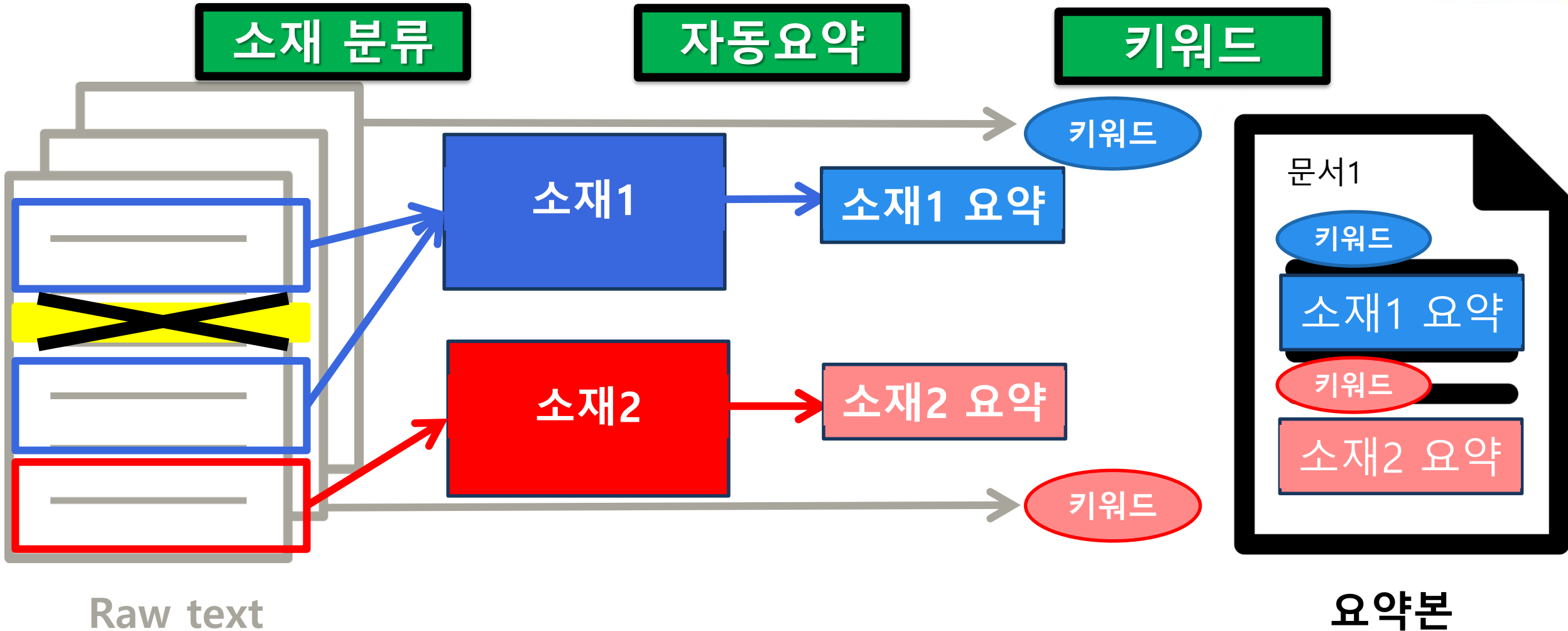


문서 자동 요약

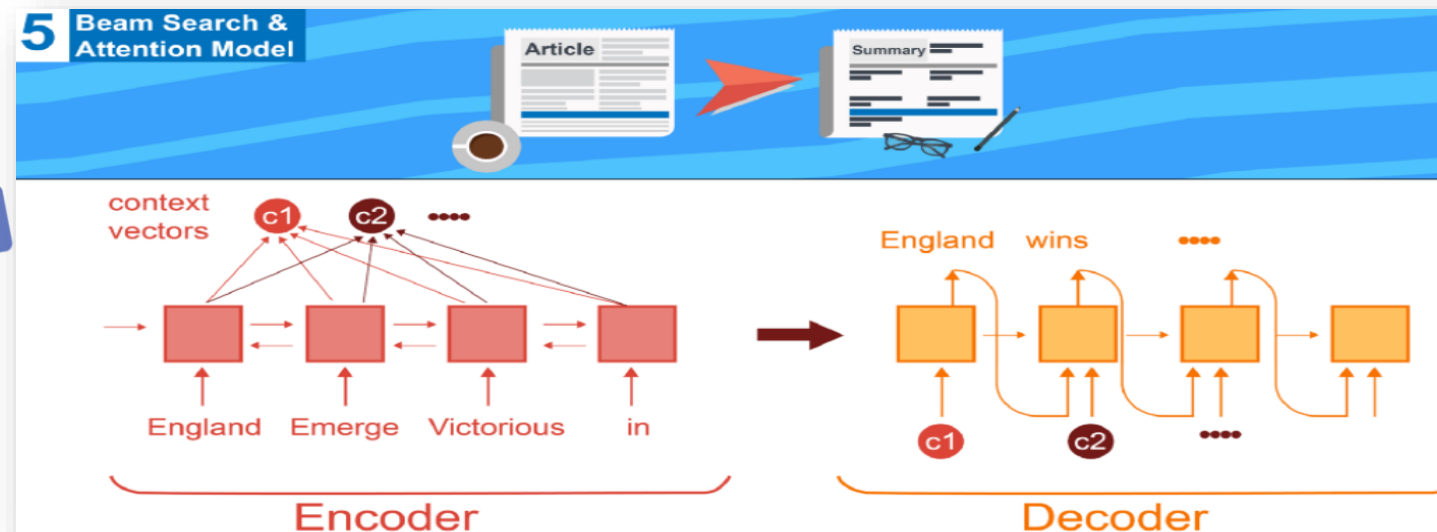
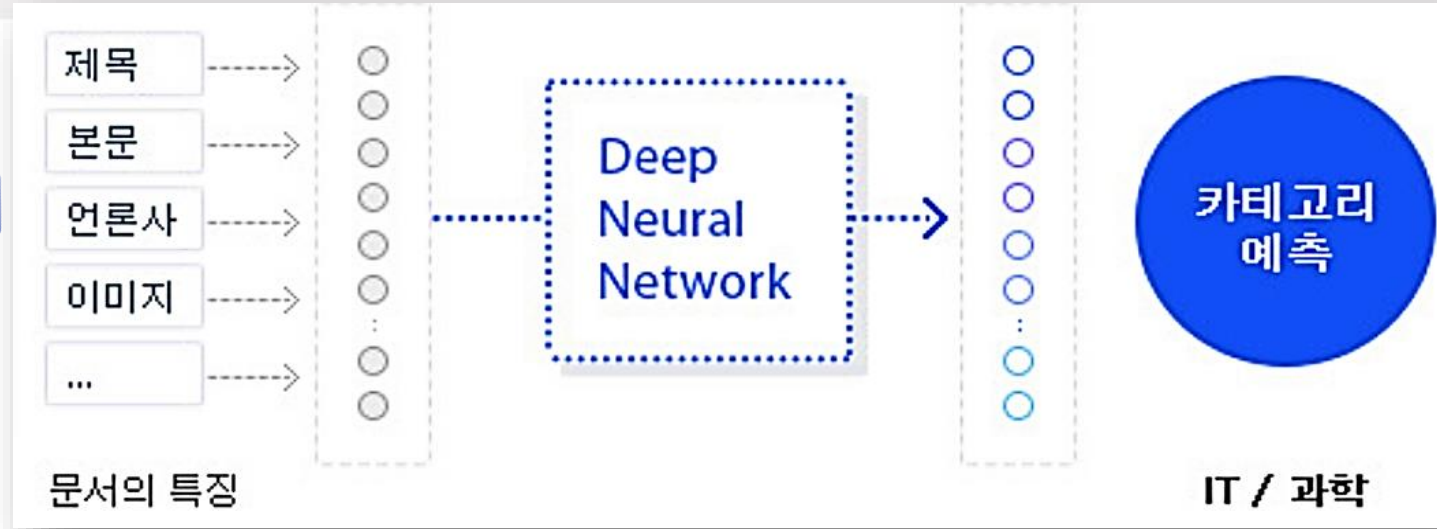
머신 러닝을 이용해 추상적 요약



서비스 소개 (사용자 측면)



서비스 소개 (기술 측면)



가치 제안



한글 전처리

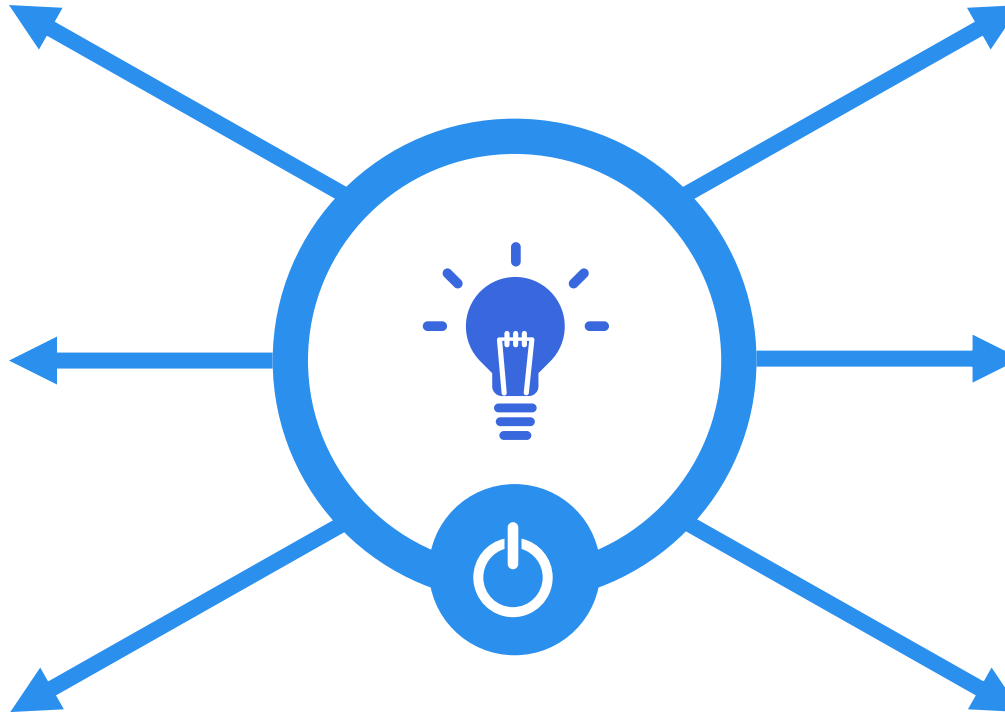
자연어 처리의 필수 과정인 전처리 과정은 머신 러닝의 기초가 됨
어절로 모든 단어가 나뉘는 영어와 달리 **어절로 분석되지 않는 한글**을 형태소 분석 후 토큰화하여 단어 임베딩 해줌

키워드 추출

전체 텍스트에서 가장 중요한 키워드를 추출하여 요약 본에서 내용의 한눈에 보게끔 도와줌

소재 분류

여러 주제의 문서들이 함께 텍스트로 입력 시, 머신 러닝을 이용한 소재 분류를 통해 1차적으로 문단 분류 기준이 됨



추상적 요약

상용화 되어 있는 추출적 요약, 가독성이 떨어지는 랭킹알고리즘이 아닌 아직 개발 중인 추상적 요약 기술 사용
마치 인간이 요약한 듯한 결과를 도출

요약본 자동 생성

체계적 요약 과정을 통해 여러 문서들을 함께 입력하였을 때도 소재 별 키워드, 요약을 시각적으로 구조화하여 요약본을 생성함

높은 적용성

머신 러닝을 이용한 '요약' 기술은 사용될 시장이 많음
효율성을 중요시하는 시대에, 요약 된 시각적 텍스트는 경제성이 높음

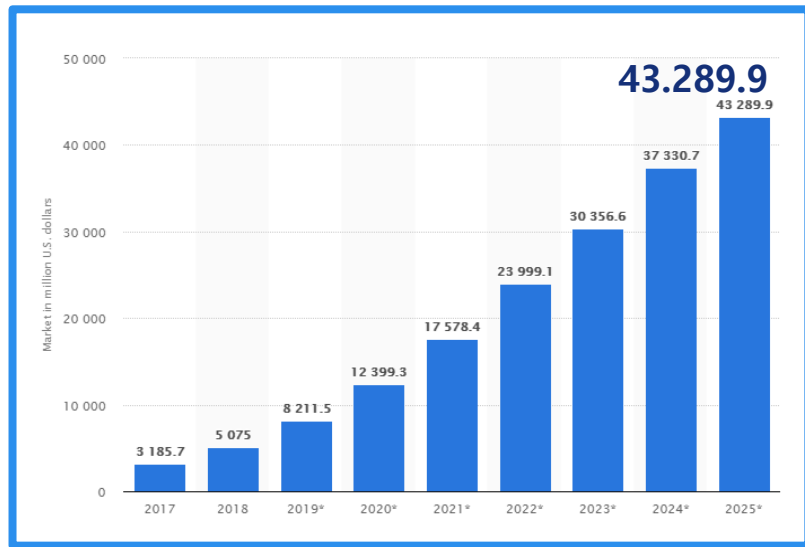


4. 시장 분석

시장 현황 및 전망



세계 NLP기술 시장 규모(2025)



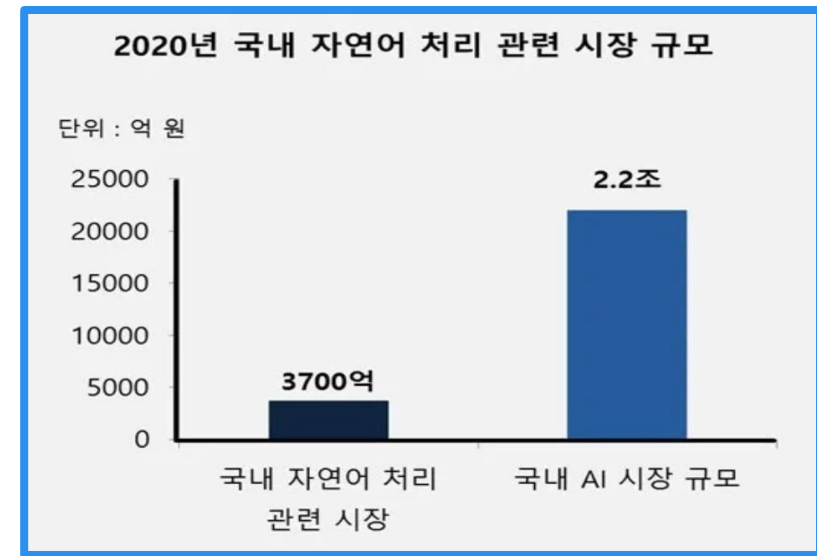
□□: statista.com (2019.7.11)



약 432억 달러

Statista가 발표한 글로벌 자연어 처리 시장 규모는 2017년 약 32억 달러에서 2025년 약 432억 달러로 확대될 전망

국내 NLP기술 시장 규모(2020)



□□: □□□□□



약 3700억원

2020년 국내 인공지능 시장은 2.2조 규모로, 자연어 처리 연관 시장규모는 3,700억 원 규모로 성장할 것으로 전망

시장 현황 및 전망



□□□□ (□□□)	- 자연어 번역의 퀄리티를 높이기 위한 빅데이터 기반의 Corpus(말뭉치) 수집과 머신러닝(Machine Learning)을 이용한 번역품질 개선기술	머신러닝
	- 자연어 대화형 서비스의 출현과 음성인식 정확도 개선을 위한 자동음성인식기술의 진보	자동음성기술
	-NLP 공개 플랫폼을 통한 생태계 구성과 이를 둘러싼 글로벌 플랫폼 사업자 간 경쟁 심화	대화형플랫폼

□: □□□□ □□□ □□
(□□: platum □□□□-AI□□□ □□ □□□□ □□□□ □□□)

기술 로드맵



경쟁업체 분석



□ □ □ □ □ □

- ✓ AI 기반 자동 추출 요약 기술
- ✓ 최대 3문장까지 요약
- ✓ 네이버 뉴스면의 뉴스면의 정치.경제.사회.IT.생활.세계.랭킹 분야에 적용

향후 요약 품질을 높이고, 라디오 대담 또는 인터뷰에 추가 적용 예정



SUMEET(□ □)

- ✓ abstractive summarization
- ✓ Sequence to sequence 방식
- ✓ 문서 내용 요약 및 키워드 추출

한국어 abstractive summarization에 도전



□ □ □ □ □

- ✓ abstractive summarization
- ✓ Text summarization with TensorFlow
- ✓ 뉴스 헤드라인 만들기
- ✓ Sequence to sequence 방식

전체내용이 아닌 서두 일부 내용을 바탕으로 문장을 만듦 (더욱 긴 글 요약을 연구중)

SWOT분석 및 전략



<div>□□□□</div> <div>□□□□</div>		□□(O)	□□(T)
		<ul style="list-style-type: none"> 정부의 AI와 NLP기술 사업 지원 NLP기술, 자동 요약 기술 시장의 지속적인 성장 가능성과 그에 따른 새로운 시장 선점 기회 	<ul style="list-style-type: none"> 편집권 침해 문제 새로운 경쟁사 출현
□□(S)	<ul style="list-style-type: none"> 비싼 하드웨어 사용하지 않아 초기 비용 부담이 적음 현 한국 시장에 상용화가 덜 이루어진 추상 요약 기술 사용 	<ul style="list-style-type: none"> 저비용과 기술의 차별화를 추구 정부의 AI 사업화 비용과 기술 개발 집중 지원.이를 위해 AI사업 기반 스타트업, 벤처 기업대상 자유 공모 추진(2019년 26.25억원/최대 2년간 지원 예정) 	<ul style="list-style-type: none"> 추상 요약 기술로 경쟁력을 높임
□□(W)	<ul style="list-style-type: none"> 한국어 위주의 시스템으로 해외시장 진출 한계 언어의 특성에 따른 성능의 한계 신기술 신시장 창출에 따른 위험 	<ul style="list-style-type: none"> 블루오션 전략으로 경쟁없이 새로운 시장 진입 및 기술의 가치와 비용을 동시 추구 	<ul style="list-style-type: none"> 추후 다양한 언어를 지원하도록 함 요약 품질을 높여 향후 다양한 분야의 요약에 사용되도록 함 정보의 수집.보관.저장 활용에 대한 대응 방안 마련

SO ST
WO WT



5. 제품 수익 모델

마케팅 전략(STP전략)



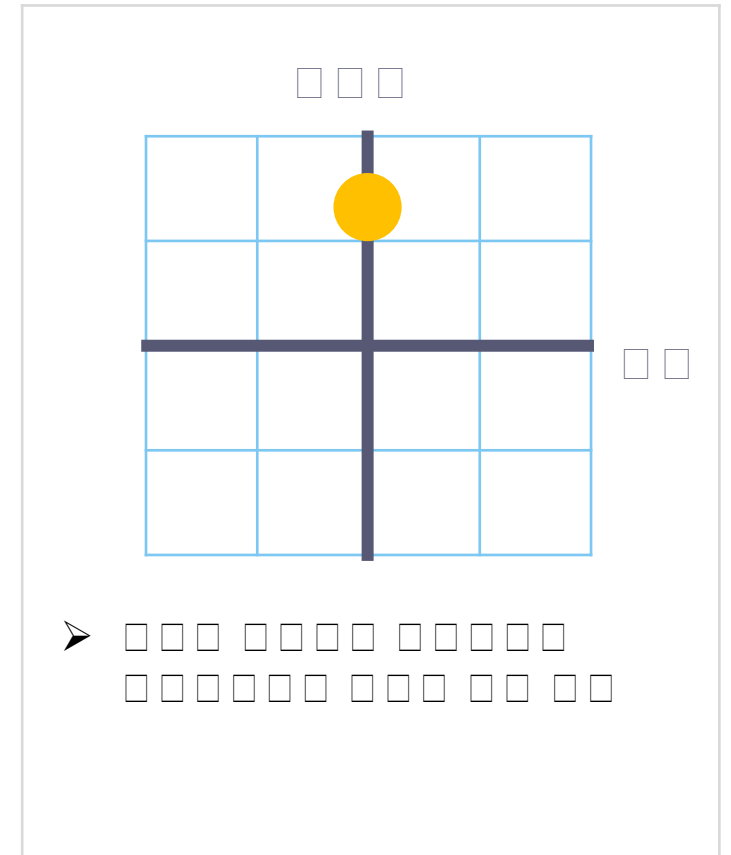
시장세분화

인구 통계학적 분류	10대에서 60대까지 실생활이나 인터넷에서 정보를 자주 접하는 연령대
심리적 분류	빠른 시간 내 정확하고 요약된 다량의 정보를 얻고 싶어하는 사회인. 많은 문서를 접하거나 막대한 정보량으로 인해 피로를 느끼는 사람들
경제적 분류	많은 것을 습득해야하는 학생이나 경제활동을 하고 있는 사회인, 웹 활동에 많은 하루의 시간을 할애하는 대상

타겟팅

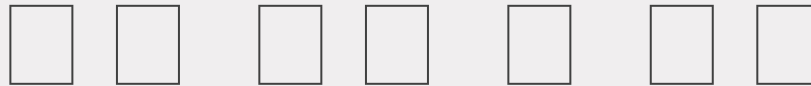
타겟팅	<ul style="list-style-type: none"> 매일 아침 긴 뉴스를 볼 시간 없어 최근 사회가 어떻게 흘러가는지 알 수 없는 30대 직장인 습득할 지식과 공부할 양이 많은 17세 고등학생 빨리 변화하는 정보와 트렌드의 선두주자로 있으며 웹 서핑을 즐기는 20대 대학생 너무나 많은 정보와 빨리 변화하는 세상 속에서 서서히 뒤쳐지는 60대 회사에서 대량의 문서 정리에 지친 40대 회사원
-----	---

포지셔닝(목표)





6. 향후 관리 및 계획



□ □ □ □ □ 多
- □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □
□ □ test □ □ □



- □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □
□ □ □ □ □ (□ □ □ □ □
□ □) □ □ □ □ test □ □ □

2 Text Representation



Text Representation for

Abstractive	Text	Summarization	using	GO	google-colab
0	1	2	3	4	5

□ □ □ □ □ □ □ □ □ □ , □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □



□ □ □ □ □ □ □ □
- □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □
□ □



□ □ □ → □ □ □
- □ □ □ □ □ □ □ □ □ □
□ □ □ , STT □ □ □
□ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □ □ □ □ □
□ □



□ □ □ □ □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ → □ □ □ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □



kt ds

AICentro (에이아이센트로)



7. 업무 분장

팀 업무 분담



이름	역할	앞으로의 유지보수	비고
박주영	Attention mechanism 모델 설계 및 테스트, Seq2seq(Neural Network) 모델 설계 및 테스트	Attention mechanism 관련 논문 참고, attention mechanism의 계산에 관한 기술 정확하게 이해, Seq2seq 관련 논문 참고하여 가장 좋은 모델 설계	팀장
김아연	텍스트 토픽 별 분류/추출적 요약 TEST	텍스트 토픽 별 분류(LSTM 또는 문서 유사성을 이용할 계획), 추출 적 요약을 위해 Lexrank TEST할 계획	팀원
김혜원	텍스트 전처리 과정	형태소 토큰화(현재 있는 형태소 분석기 수정), 품사 tagging를 위한 형태소 분석기 선택, 불용어 단어 제거	팀원
홍승환	텍스트 전처리 과정	형태소 토큰화(현재 있는 형태소 분석기 수정), OOV(Out of vocabulary)를 위한 BPE 알고리즘과 WPM알고리즘 이해	팀원
이산가 비두사	RNN을 이용하여 TEXT 생성 TEST, Seq2seq(Neural Network) 모델 설계 및 테스트	RNN을 이용하여 텍스트 생성 TEST, chat-bot TEST, neural machine translation 모델로 알려진 google의 seq2seq 참고	팀원



Thank You

□ □ □ □ □ .

