

특허 문서 요약 웹 서비스

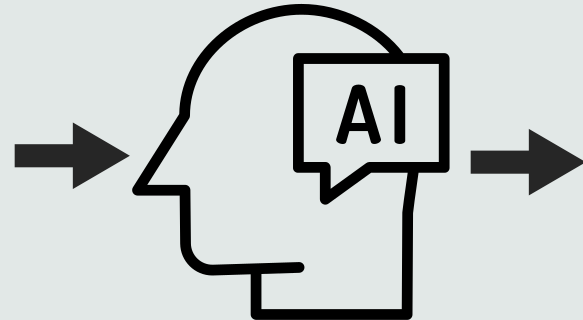
8조 팀장 심민기
김천구
이동훈
박종민

순서

1. 수행 배경 및 목표
2. 시스템 요구분석 및 정의
3. 시스템 설계 (구조 및 동작)
4. 진행상황
5. 이슈사항 및 해결방안
6. 향후 일정

수행배경 및 목표

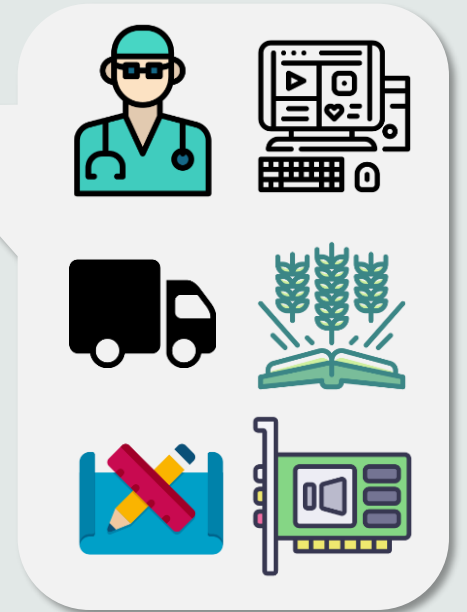
기존 특허 문서의 구조



요약된 문서 제공



다양한 분야 도움



- 대부분 특허 요약문은 청구항을 그대로 삽입
- 해당 요약문만으론 이해가 어려움

시스템 요구분석 및 정의

회사 측 요구 분석

기존 요구 사항

웹서비스 제공

- 기존 회사 서비스 '바로날인'의 부가서비스 제공
- Spring을 통한 'One Page Web'으로 구현 요구

추가 요구 사항

AI API 제공

- Spring 서버를 활용하여 HTTP API 구성(BM 활용)
- 인증 토큰을 활용한 선택적 결과 반환 대응

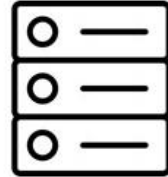
시스템 요구분석 및 정의

전체 프로세스 정의

특허 문서 요약 프로세스



특허 PDF



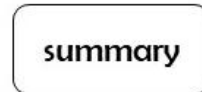
서버 업로드



PDF 내용 추출



내용 요약 AI



요약 결과 반환

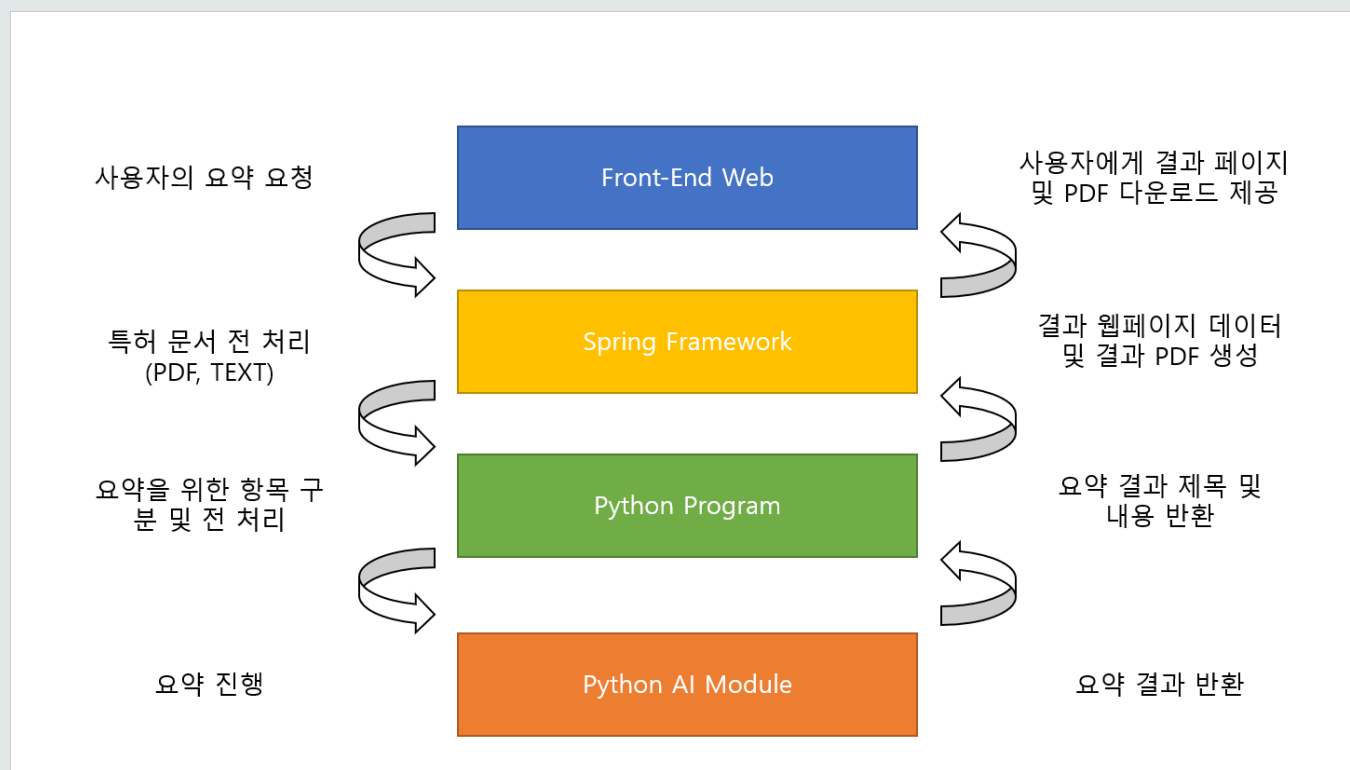


특허 요약 결과 페이지 및 PDF 다운로드 제공



시스템 설계 (구조 및 동작)

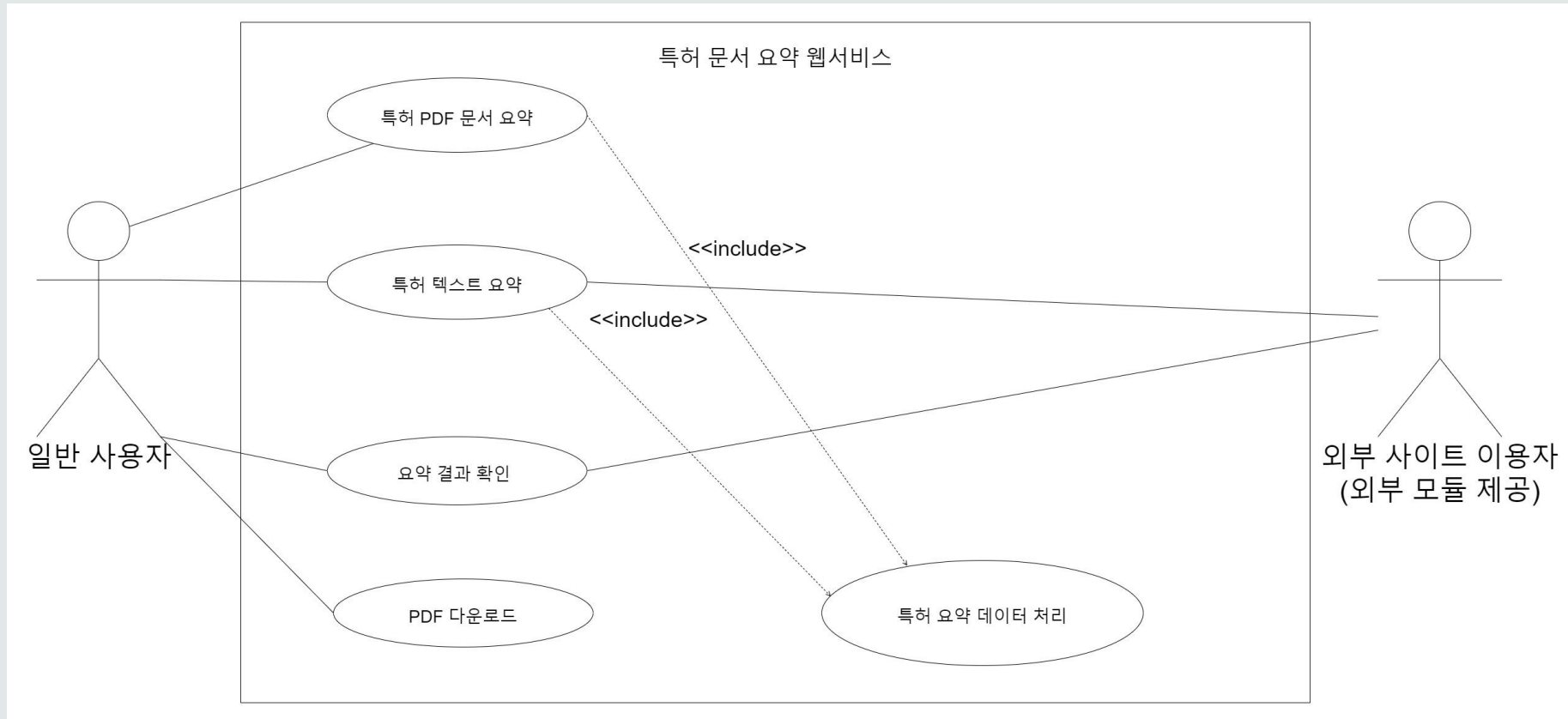
전체 프로세스의 다이어그램화



전체 시스템의 계층 분석

시스템 설계 (구조 및 동작)

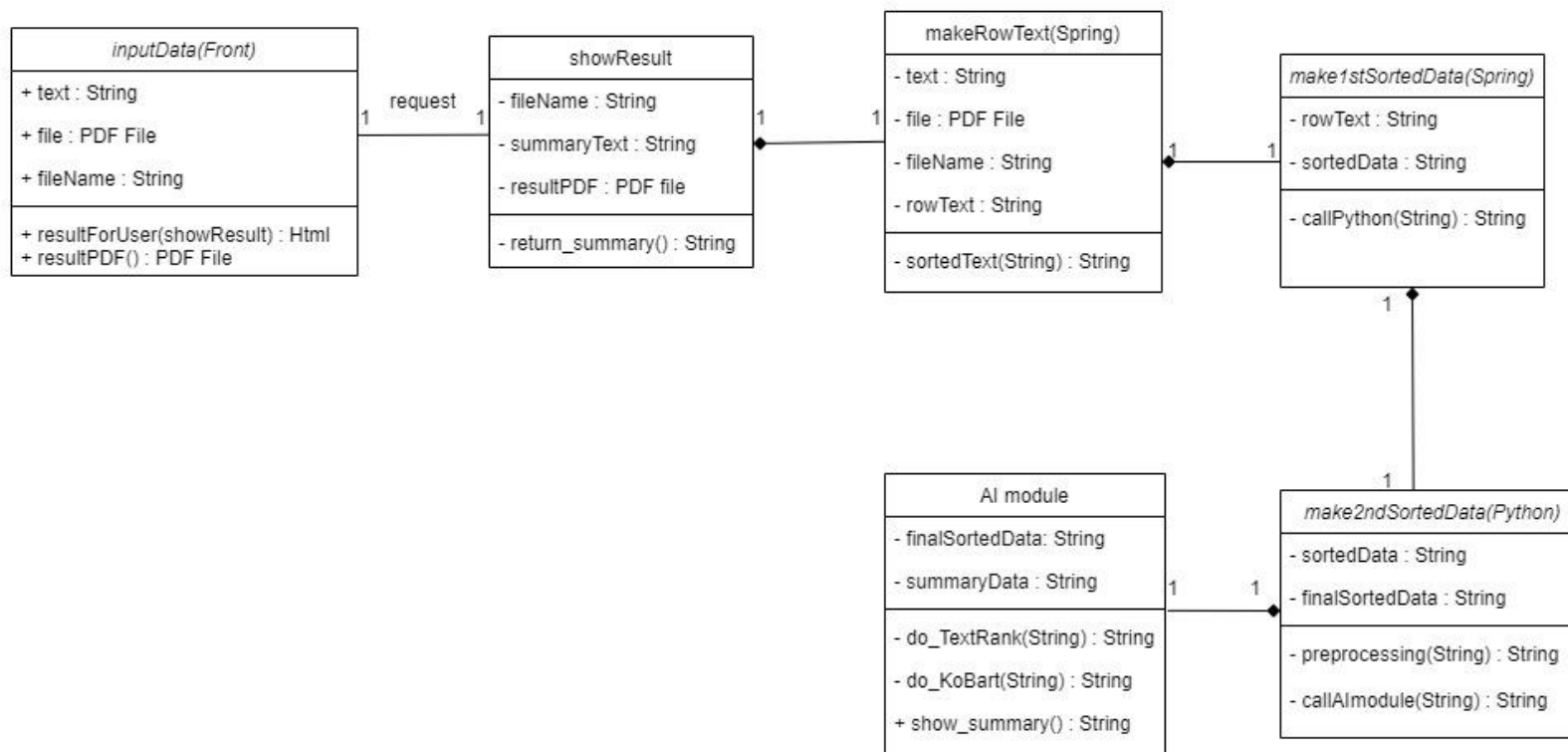
전체 프로세스의 다이어그램화



USE CASE Diagram

시스템 설계 (구조 및 동작)

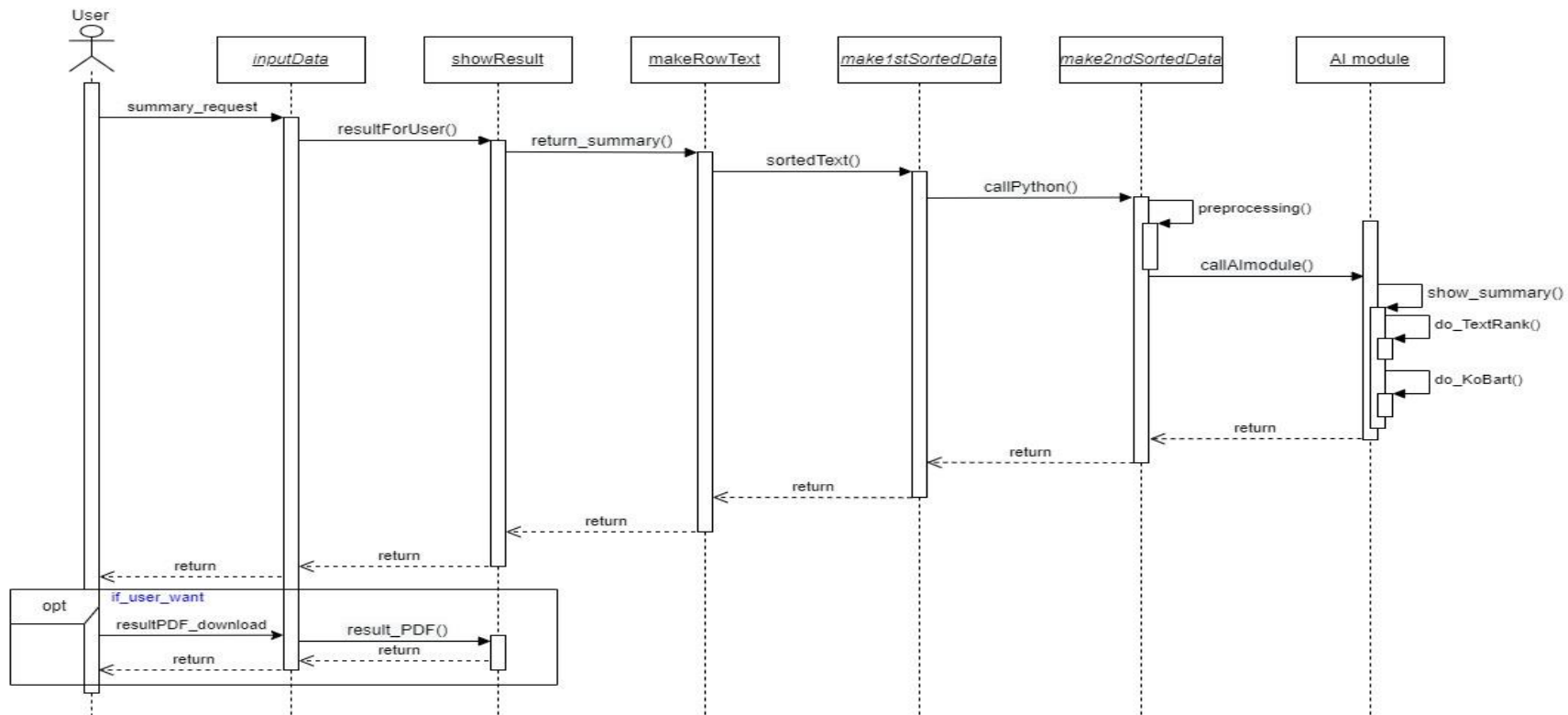
전체 프로세스의 다이어그램화



CLASS Diagram(구조)

시스템 설계 (구조 및 동작)

전체 프로세스의 다이어그램화



SEQUENCE Diagram(동작)

진행 상황

★ 는 리스크 및 요구 발생 대응

[illegible]

진행 상황

수상 내역(AI Hub 학습데이터 활용 사업화 아이디어 해커톤)

Smart Media
한국스마트미디어학회

AI Hub 학습데이터 활용 사업화 아이디어 해커톤 개최

(사)한국스마트미디어학회에서는 AI Hub 학습데이터를 활용한 사업화 아이디어 해커톤을 개최하여, 데이터 사용자 유형의 확대와 창의적 지능 서비스 발굴을 통해 국가 AI/SW 역량 강화에 기여하고자 함

1. 대회배경

- (사)한국스마트미디어학회는 스마트미디어 과학 및 공학 분야의 전문가로서 각자의 전문기술 연구, 개발을 통해 인류의 행복과 복지 및 안전에 공헌하도록 노력하고 있음
- 급변하는 현대 사회의 지능미디어 기술을 발전시키고 미래사회를 준비하기 위해서는 빅데이터 기반의 인공지능 기술의 개발이 무엇보다도 중요하다는 인식 속에 AI Hub 학습데이터를 활용한 사업화 아이디어 발굴이 무엇보다도 중요함
- 한국전통정보사회진흥원은 “AI 학습용 데이터 구축 지원사업”을 통해 미래 국가·사회 전반의 혁신을 좌우할 AI 강국으로 도약하기 위한 핵심 기반인 대규모 AI 학습데이터를 구축하고 있음



- 특히, 수집된 데이터를 활용하기 위해 AI Hub에서 'AI 컴퓨팅', 'AI 바우처', 'AI S/W'를 제공하고 있으며, 활용사례를 제시하여 AI 허브의 데이터로 개발된 서비스를 직접 체험해볼 수 있어, 사용자는 데이터의 활용 방법과 기술을 터득할 수 있고 활용 아이디어를 통해 다양한 신규 및 2차 서비스 확대를 기대하고 있음
- 이번에 구축되는 AI Hub 플랫폼을 통해 다양한 AI 서비스가 개발될 것을 기대하며, (사)한국스마트미디어학회에서는 사용자 유형의 확대와 창의적 지능 서비스 발굴을 위해, AI Hub 학습데이터를 활용한 사업화 아이디어 해커톤을 개최함

상 장

아텍 은상

팀 명 : AI꿈나무

팀 원 : 심민기 김천구 이동훈 박종민

위 참가자는 한국스마트미디어학회에서 진행한 AI Hub 학습데이터 활용 사업화 아이디어 해커톤 행사에서 우수한 성적으로 입상하였기에 이 상장을 드립니다.

2022. 04. 22

사단법인 한국스마트미디어학회장 최



진행 상황

AI

- AI 모델 결정 : KoBART
- KoBART 모델 환경 구축 및 실습
- 데이터 셋에 맞게 학습 및 파인튜닝

Data

- Dacon 한국어 문서 생성요약 AI 경진대회 의 학습 데이터를 활용함
- 학습 데이터에서 임의로 Train / Test 데이터를 생성함
- 데이터 탐색에 용이하게 csv 형태로 데이터를 변환함
- Data 구조
 - Train Data : 34,242
 - Test Data : 8,501
- default로 data/train.tsv, data/test.tsv 형태로 저장함

news	summary
뉴스원문	요약문

- 참조 데이터
 - AIHUB 문서 요약 데이터 (<https://aihub.or.kr/aidata/8054>)

```
import torch
from transformers import PreTrainedTokenizerFast
from transformers import BartForConditionalGeneration
```

```
tokenizer = PreTrainedTokenizerFast.from_pretrained('digit82/kobart-summarization')
model = BartForConditionalGeneration.from_pretrained('digit82/kobart-summarization')
```

```
text = ""
```

1일 오후 9시까지 최소 20만3220명이 코로나19에 신규 확진됐다. 또다시 동시간대 최다 기록으로, 사상 처음 20만명대에 준 방역 당국과 서울시 등 각 지방자치단체에 따르면 이날 0시부터 오후 9시까지 전국 신규 확진자는 총 20만3220명으로 집계됐다. 국내 신규 확진자 수가 20만명대를 넘어선 것은 이번이 처음이다.

동시간대 최다 기록은 지난 23일 오후 9시 기준 16만1389명이었는데, 이를 무려 4만1831명이나 웃돌았다. 전날 같은 시간 확진자 폭증은 3시간 전인 오후 6시 집계에서도 예견됐다.

오후 6시까지 최소 17만8603명이 신규 확진돼 동시간대 최다 기록(24일 13만8419명)을 갈아치운 데 이어 이미 직전 0시 기준 17개 지자체별로 보면 서울 4만6938명, 경기 6만7322명, 인천 1만985명 등 수도권이 12만5245명으로 전체의 61.6%를 차지했다. 비수도권에서는 7만7975명(38.3%)이 발생했다. 제주를 제외한 나머지 지역에서 모두 동시간대 최다를 새로 썼다.

부산 1만890명, 경남 9909명, 대구 6900명, 경북 6977명, 충남 5900명, 대전 5292명, 전북 5150명, 울산 5141명, 광주 5099명, 전남 4999명, 강원 4999명, 충북 4999명, 제주 4999명 등 집계를 마감하는 자정까지 시간이 남아있는 만큼 2일 0시 기준으로 발표될 신규 확진자 수는 이보다 더 늘어날 수 있다. 이 한편 전날 하루 선별진료소에서 이뤄진 검사는 70만8763건으로 검사 양성률은 40.5%다. 양성률이 40%를 넘은 것은 이번이 처음이다. 이날 0시 기준 신규 확진자는 13만8993명이었다. 이를 연속 13만명대를 이어갔다.

```
text = text.replace('\n', ' ')
```

```
raw_input_ids = tokenizer.encode(text)
input_ids = [tokenizer.bos_token_id] + raw_input_ids + [tokenizer.eos_token_id]
```

```
summary_ids = model.generate(torch.tensor([input_ids]), num_beams=4, max_length=512, eos_token_id=1)
tokenizer.decode(summary_ids.squeeze().tolist(), skip_special_tokens=True)
```

```
'1일 0 9시까지 최소 20만3220명이 코로나19에 신규 확진되어 역대 최다 기록을 갈아치웠다.'
```

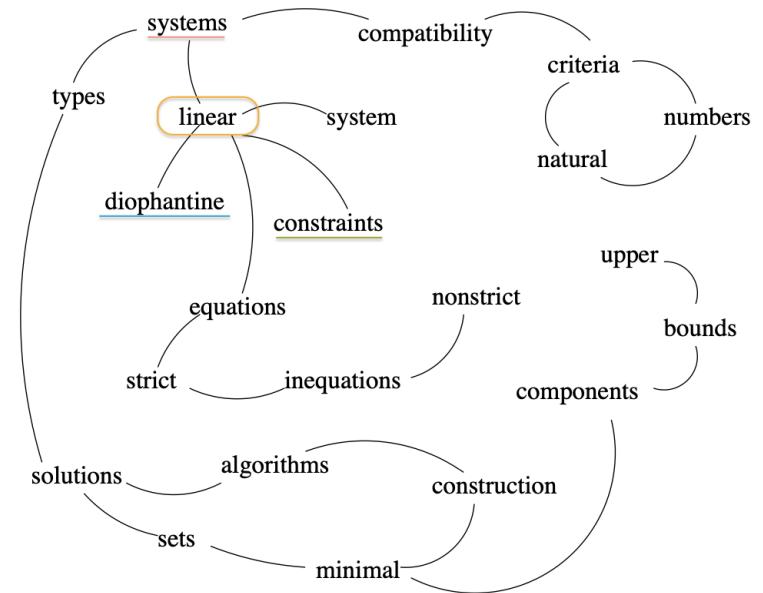
진행 상황

AI

- 추가적인 알고리즘 : TextRank
- TextRank 알고리즘으로 추출요약 진행 후, 튜닝한 KoBART모델을 사용해 생성요약을 진행함

Keyword Extraction

Compatibility of systems of linear constraints over the set of natural numbers. Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and nonstrict inequations are considered. Upper bounds for components of a minimal set of solutions and algorithms of construction of minimal generating sets of solutions for all types of systems are given. These criteria and the corresponding algorithms for constructing a minimal supporting set of solutions can be used in solving all the considered types systems and systems of mixed types.



이슈사항 및 해결방안

- 이슈사항

회사 측 웹 페이지 디자인 추가 제안 발생

- 해결방안

- 높지 않은 Risk로 회사 측 의견 수용하여 웹페이지 형태를 일부 수정

- >> Background Color 제공으로 가독성 향상

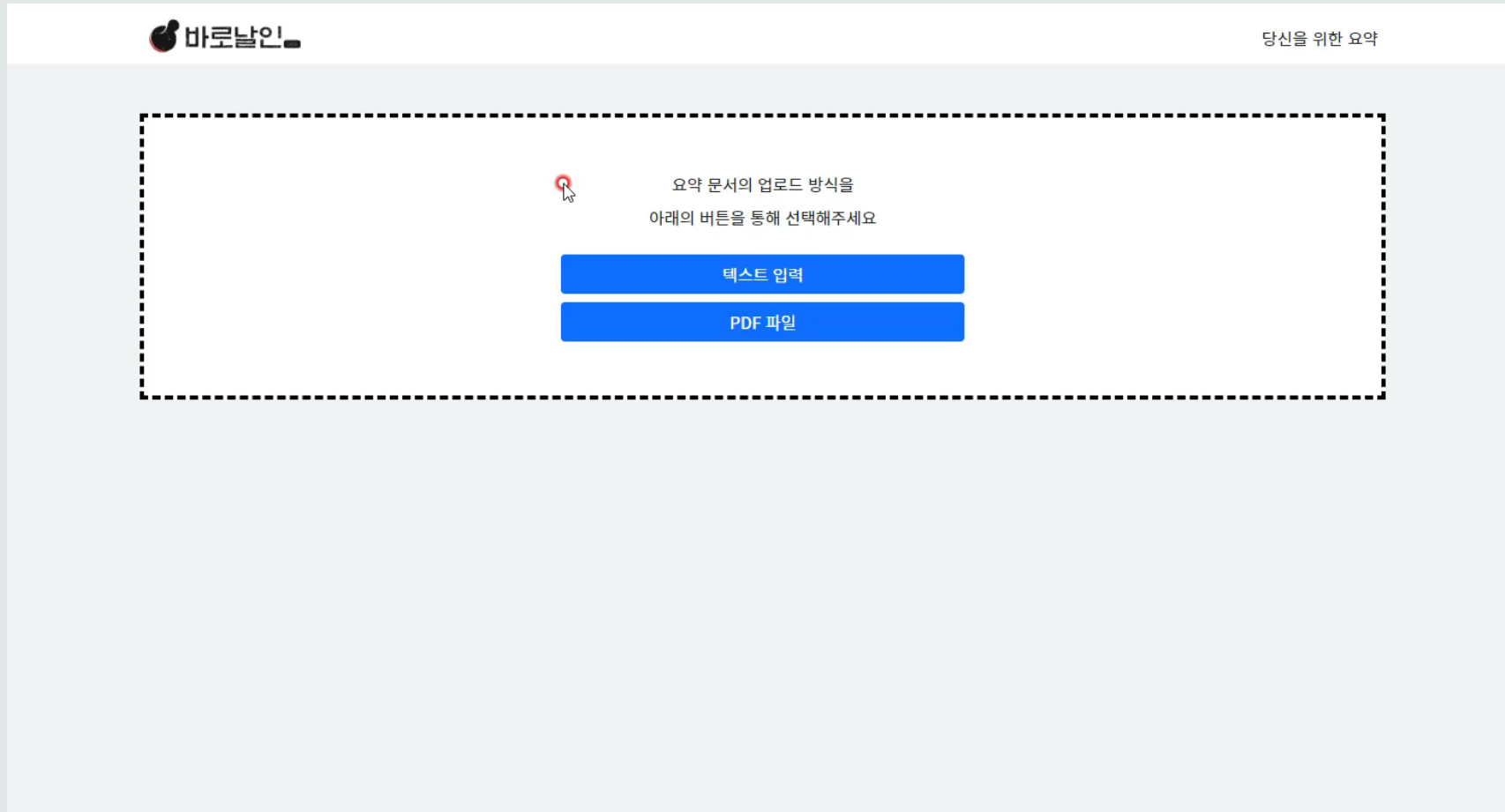
- >> 원문 / 요약문 모두 제공하는 방식으로 변경하여 요약의 정도 파악할 수 있도록 제공

- >> 로딩 gif 제공으로 대기 상태를 사용자가 인식하도록 구성

- >> 전체적인 비율 조정으로 모바일/1200px 태블릿/1920 PC 에서 모두 반응할 수 있도록 구성.

이슈사항 및 해결방안

- 변경된 모습



향후 일정

간트 차트를 통한 향후 일정 분석

★ 는 리스크 및 요구 발생 대응

WBS 번호	작업 제목	작업 책임자	시작일	마감일	기간	작업 완료 비유	4분기										5분기																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
							0주(5/2~					11주(5/9~					12주(5/16~					13주(5/23~					14주(5/30~					15주(6/6~																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
							목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금	월	화	수	목	금																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
3	추가 개발 및 알파 테스트																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													

- 테스트 배포
- AI 품질 향상을 위한 추가 활동(연구실 협업)

향후 일정

- 2022 한국컴퓨터종합학술대회 (KCC 2022) 발표
 - 논문접수마감: 2022-04-22(금)
 - 심사결과발표: 2022-05-27(금)

KCC2022

6월 29일(수)~7월 1일(금), ICC 제주
"디지털 혁신을 이끄는 소프트웨어"

[홈페이지 바로가기](#)

// 초대의 말씀

한국정보과학회는 회원들의 연구 성과를 발표하고 학술 정보를 나눔과 동시에 회원 상호 간의 친목을 도모할 수 있는 기회를 마련하고자, 정기적으로 한국컴퓨터종합학술대회(KCC)를 개최하고 있습니다. KCC2022는 6월 29일(수)부터 7월 1일(금)까지 제주도 ICC(국제컨벤션센터)에서 개최될 예정입니다.

KCC2022는 "디지털 혁신을 이끄는 소프트웨어"라는 주제로, 소프트웨어가 선도하는 디지털 혁신을 통해 변화되는 세상을 준비하는 기회를 제공하고자 합니다.

KCC2022에서는 논문 발표, 튜토리얼, 특별세션 등의 학회 주관 행사와 더불어 학술 분과 주관 워크샵 및 협력 워크샵 등 다양한 산학연 학술 행사를 진행할 예정입니다. 이번 학술발표회에서는 채택 논문 상위 10% 내외의 우수 논문과 발표 논문 상위 10% 내외의 우수 발표 논문을 선정하여 학회 논문지에 게재를 추천하고, 학부/대학원생들의 참신한 아이디어 및 소프트웨어 개발 능력 향상을 위해 산업계와 함께 SW 경진대회를 진행할 계획입니다.

마지막으로 이번 행사의 성공적 개최를 위해 노력해 주시는 실무적 회장님과 학회 임원 여러분, 프로그램위원회의 조직위원회의 위원님들, 정보과학 발전에 항상 협력해 주시는 유관 기관 및 산업계 관계자 여러분, 유용한 지식을 공유해주신 모든 발표자 여러분께 깊은 감사의 말씀을 드립니다. 본 학술대회가 우리 학회 구성원 모두에게 유익한 행사가 될 수 있도록 다시 한번 회원 여러분의 적극적인 참여와 협조를 부탁드립니다.

2022년 3월
Korea Computer Congress 2022

대 회 장 재진석(인원대)
프로그램 위 원 장 문양세(강원대)
조 직 위 원 장 김성백(제주대)
프로그램부위원장 권혁윤(서울과기대), 박영호(숙명여대), 서동민(KIST), 이종욱(성균관대),
임성수(충남대), 임한승(강원대), 주재걸(KAIST)

// 논문 모집

- 논문내용: 정보과학에 관한 학술논문 및 기술보고 등
- 모집분야: 컴퓨터 공학 전 분야
- 논문분량: 2~3쪽 (작성양식)
- 제출방법: 홈페이지 온라인 접수 (바로가기)
- 논문접수마감: 2022년 4월 22일(금)
- 심사결과발표: 2022년 5월 27일(금)
- 논문최종분결: 2022년 6월 7일(화)
- 우수논문은 정보과학회 논문지 게재 추천

// 튜토리얼 제안 모집

- 모집분야: 정보과학 분야 다수의 관심 대상 자유 주제
- 제출내용: 강의제안서 및 연사약력 (작성양식)
- 모집마감: 2022년 4월 15일(금)
- 문의 및 제출: 권혁윤 부위원장(서울과기대 교수) [✉](#)
- ※ 프로그램위원회 심의를 통해 채택되지 않을 수도 있음

// 협력워크샵 제안 모집

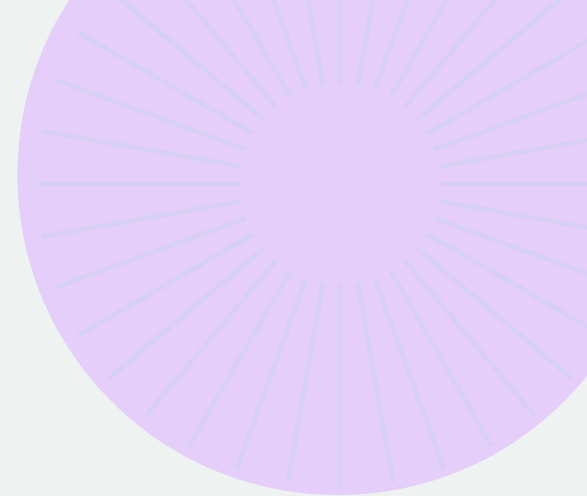
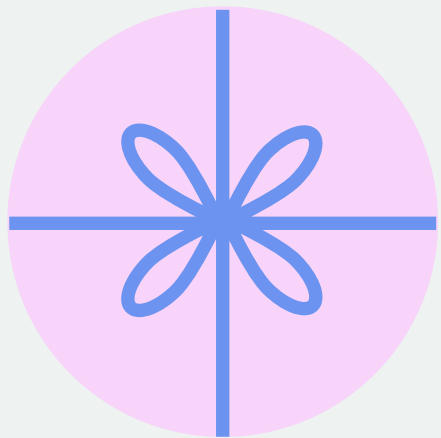
- 모집내용: 세미나, 사업단, 과제 성과발표회 및 워크샵
- 제출내용: 워크샵 제안서 (작성양식)
- 모집마감: 2022년 4월 15일(금)
- 문의 및 제출: 서동민 부위원장(KISTI 책임연구원) [✉](#)
- ※ 프로그램별 마감일정은 사정에 따라 다소 변경될 수 있습니다.
- ※ 자세한 사항은 홈페이지를 참고하시기 바랍니다.

// 신진연구자 초청

- 프로그램: 최신 연구 내용 소개 및 네트워킹
- 초청대상: 만 39세 이하이거나 박사학위 취득 후 7년 이내
- 참가신청: 설명, 소속, 연락처, 연구분야 기재하여 신청
- 문의 및 신청: 이종욱 부위원장(성균관대 교수) [✉](#)

대회장 재진석(인원대) 프로그램위원장 문양세(강원대) 조직위원장 김성백(제주대)
부위원장 권혁윤(서울과기대), 박영호(숙명여대), 서동민(KIST), 이종욱(성균관대), 임성수(충남대), 임한승(강원대), 주재걸(KAIST)

홈페이지: <https://www.kisee.or.kr/conference/kcc/2022/> 문의: 박보혜 주임(정보과학회) bhpark@kisee.or.kr 02-588-9247



발표를 마치겠습니다.

감사합니다.

