

인공지능 학습용 데이터 구축 사업

문서요약 텍스트 AI 데이터

1. 과제 목표

중소·벤처, 스타트업 등 민간의 인공지능 기술개발 촉진,
인공지능산업 육성을 위해 대규모 인공지능 학습용 데이터
구축·개방

2. 배경 및 필요성

□ 추진배경

- AI가 스스로 인식, 이해할 수 있으려면 AI SW가 사물간 연관성을 이해할 수 있는 형태로 가공된 대규모 AI 학습용 데이터 확보가 중요
- 미국, 유럽 등 AI 선도국에서는 AI 기술 발전을 위해 정부와 민간 협업·투자로 대규모의 데이터를 대량으로 구축하여 공개

□ 필요성

- AI 선도들은 인공지능(AI) 등 4차 산업혁명 주도기술을 활용해 미래 산업 변화의 주도권을 잡기위한 국가차원의 지원 적극 추진 중
- 반면 국내 중소·벤처기업들은 AI 학습용 데이터를 자체 구축하기에 많은 시간과 비용이 소요되고 원천 데이터 확보의 어려움 호소

No	공모 분야명
1	(지정공모) 문서 요약 텍스트 AI데이터
2	(지정공모) 대용량 동영상 콘텐츠 AI데이터
3	(지정공모) 딥페이크 변조영상 AI데이터
4	(지정공모) 수어(手語)영상 AI데이터
5	(지정공모) 시각정보기반 질의응답 AI데이터
6	(지정공모) 전문분야 한·영 말뭉치 AI데이터
7	(지정공모) 랜드마크 이미지 AI데이터
8	(지정공모) 자율주행드론 비행영상 AI데이터
9	(지정공모) 한국인 대화음성 AI데이터
10	(지정공모) 사람 인체·자세 3D AI데이터
11	(자유공모) 자유공모 과제

[과제 구분]

3. 문서요약 사업 정보

과제번호	지정공모 - 1
과제명	문서요약 텍스트 AI 데이터
주관기관	비플라이스포트
참여기관	(주)위고, 테스트웍스, 고려대학교, 주식회사 에이아이닷엠

4. 문서요약 사업 목적 및 목표

□ 사업목적

- 검증된 한국어 텍스트 요약 데이터를 통한 AI 요약 기술 경쟁력 확보
- 한국어 요약 데이터 및 모델을 통한 다양한 신사업 파생

□ 사업 추진 목표

- 다양한 한국어 원문 데이터로부터 정제된 추출 및 생성 요약문을 도출하여 검증된 한국어 문서 요약 AI 데이터셋 구축 및 배포
- 구축된 한국어 문서 요약 AI 데이터셋을 기반으로 추출/생성 요약 AI 알고리즘 개발 및 배포
- 추출/생성 요약 AI 알고리즘을 이용한 문서 요약 관련 서비스 구현 및 API 배포

5. 문서요약 사업 내용

□ 문서 요약용 데이터 수집 및 전처리

- 저작권이 해결된 신문기사(30만건), 기고문(3만건), 잡지기사(1만건), 논문 초록(3만건), 법원 판결문(3만건) 등 다양한 속성을 가진 한글 원문 40만건 이상 선정
- 문장분리, 띄어쓰기 수정, 오타자 수정, 형태소 분석을 통한 수식어 제거 등의 전처리 수행을 통한 요약문 작업자의 문서 독해 수월성 향상 지원

□ 문서 요약문 생성 및 검수

- 전처리된 원문으로부터 추출 요약문 선택 및 생성요약문 작성
- 생성요약문에 대한 전문가 검수 및 교차 검증 실시
- 최소 35만쌍 이상의 문서-추출/생성요약문 구축



문서요약 프로젝트 가이드

7. 문서요약 클라우드 워커 모집(이벤트진행중)

2020 인공지능 데이터 구축사업

**언어능력자
프로요약러
모집합니다!**



TESTWORKS

2020 인공지능 데이터 구축사업

참여 대상 및 참여 방법

국내 거주 대학생, 사회인 모두 환영

1. aiworks.co.kr 접속
2. 문서요약 프로젝트 참여
3. 포인트를 현금으로 전환

TESTWORKS

2020 인공지능 데이터 구축사업

문서요약 프로젝트란?

1. 신문기사를 읽고
2. 핵심 문장 3개 표시
3. 한 문장으로 요약
4. 완료 후 포인트 받기

TESTWORKS

2020 인공지능 데이터 구축사업

프로젝트 참여 혜택

**최대 2,000,000 포인트 획득
가능(건당 500 포인트)**

*10건 이상 참여 ▶ 5000 포인트 이상 현금 전환 가능

TESTWORKS

2020 인공지능 데이터 구축사업

문서요약 프로젝트 이벤트 3탄

한글의날 기념 특별 이벤트

**10000원
보너스 받기!**

기간: 10월7일(수) ~ 10월23일(금) 자정까지

가공 문서 당 1,500원의 혜택을 놓치지 마세요!
*검수 완료 건에 한함

참여 방법

1. aiworks.co.kr 접속
2. 문서요약 프로젝트 참여
3. 포인트를 현금으로 전환

*자세한 내용은 홈페이지를 참고하세요.



활동 내용

신문기사 읽고
 > 핵심 3개 문장 표시해서
 > 한 문장으로 요약하기

참여 혜택

문서요약 프로젝트 가공하기에서
 검수가 완료된 문서 10건 당
10,000원 추가 지급
 *이벤트 기간 동안 10건 당 반복 지급

구분	기준	이벤트
10건 검수 완료	5,000원	15,000원
20건 검수 완료	10,000원	30,000원
30건 검수 완료	15,000원	45,000원





감사합니다