

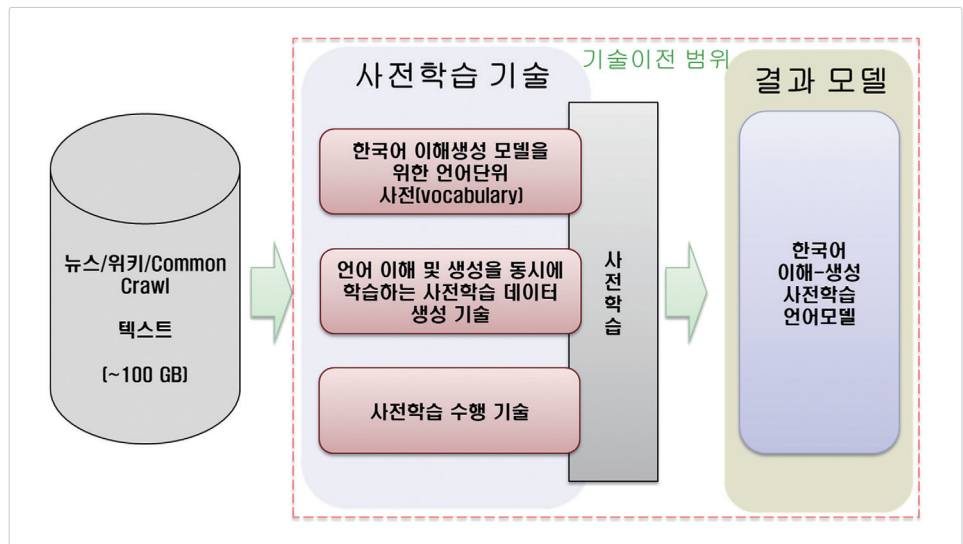
엑소브레인 한국어 이해-생성 언어모델 사전학습 기술 V1.0

기술개요

- 대용량 텍스트로부터 한국어의 구문 및 의미 정보를 학습하여, 다양한 응용 태스크(언어분석/기계독해/문서분류/요약생성/대화생성 등)에 범용적으로 활용 가능한 딥러닝 언어모델을 학습하는 기술

기술특성

- 고성능의 한국어 언어모델 학습에 필요한 언어단위 사전, 사전학습 (입력/출력) 데이터 생성, 사전학습 수행 기술 등을 포함
- Common Crawl, 위키백과, 신문기사 등, 약 95 GB(9억6천만 문장, 97억개 단어, 440억 글자)의 데이터로부터 학습한 한국어 이해생성 언어모델은 언어분석/기계독해/문서분류 등의 언어 이해 능력 뿐만이 아니라, 요약/대화 등의 언어 생성도 가능



적용분야

- 한국어 범용 도메인 질의응답
- 한국어 문장의 의미적 관계 분석
- 텍스트 빅데이터 분석 및 마이닝 기술
- 한국어 요약 / 챗봇 대화 서비스

기술완성도 (TRL)

◆ 6단계 : 파일럿 규모 시작품 제작 및 성능 평가



기술이전 내용

- ◆ 한국어 이해생성 모델을 위한 언어단위 사전(vocabulary)
- ◆ 언어 이해 및 생성을 동시에 학습하는 사전학습 데이터 생성 기술
- ◆ 사전학습 수행 기술
- ◆ 사전학습된 한국어 이해생성 언어모델

지식재산권 현황

No.	출원·등록번호	특허명	상태
1	2021-0091541	언어 이해 및 생성이 가능한 한국어 딥러닝 언어모델을 위한 사전학습 시스템 및 방법	출원

기술이전 문의

◆ ETRI 연구성과확산실 | 042-860-4881 / etri_tco@etri.re.kr