



#코퍼스 #감성대화 #감성챗봇 #우물줄개발

감성 대화 말뭉치

분야 | 한국어 용량 | 오디오, 텍스트

구축년도 : 2020 생산년월 : 2022-12 조회수 : 131,540 다운로드 : 13,202 용량 : 20.35 MB

다운로드

샘플 데이터

관심 데이터 등록

377

소개

파일 목록 (API 다운로드)

※ 내국인만 데이터 신청이 가능합니다.

문의하기

목록

데이터 개요

데이터 변경이력

버전	일자	변경내용	비고
1.2	2022-12-28	완전데이터, 라벨링데이터 재연결	
1.1	2022-12-07	완전데이터, 라벨링데이터 수정	
1.0	2021-06-18	데이터 최초 개방	

데이터 히스토리

일자	변경내용	비고
2022-12-07	어노테이션 포문 및 데이터 구조 내용 수정	

소개

크리우드 소싱 수행으로 일반인 1,500명을 대상으로 하여 음성 15,700문장 및 코퍼스 27만 문장 구축 및 세대별 감성 대화 텍스트 구축을 통해 감성 대화 엔진을 개발하여 세대별 감성 대화 서비스 제공

- 데이터 영역 : 한국어
- 데이터 유형 : 오디오, 텍스트
- 구축년도 : 2020년
- 구축량 : 10,000문장 (음성), 27만문장 (코퍼스)

구축목적

감정 인식을 위한 데이터는 크롤링이 불가능하기 때문에 직접 제작해야 하는 최소성 있는 데이터임. 60가지의 세부 감정에 대한 자연어 처리 말뭉치를 확보함으로써 다양한 AI 산업에 활용이 가능함

데이터 통계

구축 내용 및 제공 데이터량

- 음성 약 10,000 문장
- 코퍼스 27만 문장

구분	내용
데이터 구축 범위	감성 텍스트 언어 수집 -우물줄 관련 언어 의미 구조화 및 대화 응답 시나리오 동반 수집
데이터구축 규모	음성 약 10,000 파일, 코퍼스 27만 문장 수집예정 -일반인 1,500명 대상의 인터뷰 및 크리우드 소싱 수행 -우물줄 참가 대상 WQOL 대체 수집
데이터구축 일정	중 5개월 소요 예정 -데이터 설계 및 인프라 구축 1개월, 데이터 수집예정 3개월, 데이터 품질 검수 1개월

교육활동 동영상

저작도구

활용 AI 모델 및 코드

데이터 성능 지표

어노테이션 포맷 및 데이터 구조

설명서 및 활용가이드 다운로드

데이터 설명서 다운로드 구축활용가이드 다운로드

데이터 변경이력

버전	일자	변경내용	비고
1.0	2021.06.18	데이터 최초 개방	

구축목적

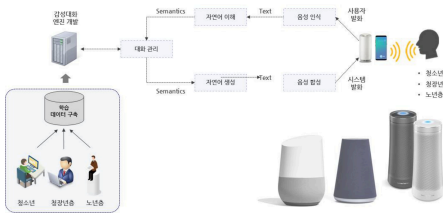
- 감정 인식을 위한 데이터는 크롤링이 불가능하기 때문에 직접 제작해야 하는 최소성 있는 데이터임. 60가지의 세부 감정에 대한 자연어 처리 말뭉치를 확보함으로써 다양한 AI 산업에 활용이 가능함

활용분야

- 인공지능 학습용 감성대화 코퍼스 데이터 구축을 통해 감성대화 엔진을 개발, 감성 챗봇 또는 음성버서 서비스 출시
- 청소년, 청장년층, 노년층 등 세대별 감성대화 코퍼스를 분리 구축, 감성대화 서비스에서도 세대별로 특화된 대화 서비스를 제공

소개

- 우물줄 관련 언어 의미 구조화 및 대화 응답 시나리오 동반한 감성 텍스트 언어 수집
- 크리우드 소싱 수행으로 일반인 1,500명 대상으로한, 음성 10,000 문장 및 코퍼스 27만 문장 구축



구축 내용 및 제공 데이터량

- 음성 약 10,000 문장
- 코퍼스 27만 문장

구분	내용
데이터 구축 범위	감성 텍스트 언어 수집 -우물줄 관련 언어 의미 구조화 및 대화 응답 시나리오 동반 수집
데이터구축 규모	음성 약 10,000 파일, 코퍼스 27만 문장 수집예정 -일반인 1,500명 대상의 인터뷰 및 크리우드 소싱 수행 -우물줄 참가 대상 WQOL 대체 수집

