

2019 공개SW 컨트리뷰톤



케라콘- 케라스 기반 챗봇 만들기

■ 프로젝트 내용

- **프로젝트명** : 케라콘 - 케라스 기반 챗봇 시스템
- **프로젝트 분야** : 인공지능 / 딥러닝
- **프로젝트 저장소** : 현재 미생성(Github repository 이용 예정)
- **활용 언어** : Python
- **프로젝트 난이도** : 중상
- **참가자 모집 유형**:
 - (필수) Python 언어에 익숙하고, 딥러닝에 대한 기본적인 이해가 있으신 분
 - 케라스 를 활용한 자연어 처리 기반 딥러닝에 관심 있는 사람
 - 협업을 통한 기술 공유를 우선시 하며, 지속적으로 성장 하는 것을 목표로 할 수 있는 사람

■ 프로젝트 개요

안녕하세요, 케라스 코리아 입니다.

이번 케라스 컨트리뷰톤의 프로젝트중 하나인 챗봇을 채택한 이유는 실생활에서 가장 쉽게 접할 수 있고, 활용 범위가 넓어 많은 사람들이 참여하고, 함께 만들어가는데 의미가 있기 때문입니다

프로젝트는 한글데이터를 기반으로 진행 할 예정이고, 먼저 자연어 처리의 기본 과정을 진행한 후 Python 환경에서 본격적인 프로젝트를 진행 하고자 합니다.

프로젝트 순서는 데이터 수집, 전처리, 라벨링, 모델 생성, 모델 평가 입니다.

영화 평점 데이터가 쉽게 구할 수 있기 때문에 이와 관련된 도메인으로 진행할 예정입니다. 팀원들의 의견과 수집 가능한 말뭉치 데이터에 따라 확장 및 변경은 가능합니다.

■ 컨트리뷰톤 가이드

Step 1: **컨트리뷰톤 기본 협업 방법 숙지**

Step 2: **개발 환경 구축**

Step 3: **자연어처리에 대한 기본 과정 이해하기**

Step 4: **챗봇 프로젝트구축진행**

Step 5: Deploy

■ 컨트리뷰톤 가이드

Step 1: 컨트리뷰톤 기본 협업 방법 숙지

- 원활한 컨트리뷰톤 진행과 올바른 오픈소스 기여를 위해 협업 방식과 틀을 숙지합니다.
- Git 을 통한 가장 기본적인 방법을 숙지합니다.
 - Github 기본적인 명령어 숙지, 기본적인 협업을 통한 프로젝트를 수행합니다.
- 팀 내에서 가장 합리적인 온라인 / 오프라인 커뮤니케이션 방법을 정의 합니다.

■ 컨트리뷰톤 가이드

Step 2: 개발 환경 구축(Keras기반 챗봇 프로젝트 환경구축)

- 개발에 가장 필요한 python 설치 및 Keras활용이 가능한 환경을 구성합니다
- 형태소 분석기는 Konlpy 패키지를 활용합니다.
- 각자 맡은 파트에서 milestone에 따라 작업을 진행합니다.
- 도메인 영역을 합의 하에 지정합니다.(잡담,특정 목적에 따른 용도 등)
- 전체 과정을 스텝 별로 나누어 서로 맡을 파트를 나눕니다
 - 데이터 수집, 전처리, 라벨링, 모델 생성, 모델 검증의 사이클을 고려
 - 한 Cycle 씩, 파트를 나눌지, 특정 단계를 위주로 할지는 협의 진행

■ 컨트리뷰톤 가이드

Step 3: 자연어처리에 대한 기본 과정 이해하기

- 프로젝트의 핵심이자, 기본인 자연어 처리에 대한 개념을 최대한 많이 숙지하도록 합니다.
→ 단어의 의미에 대해 기본적인 내용을 숙지하도록 합니다.(시소러스, 통계기반, 추론기반)
- 자연어 처리 기술적인 부분에 대해 최대한 이해합니다.
→ 기본 word2vec, 형태소 분석기 활용방법, Tokenizer, Skip gram/ CBOW 등등
- 딥러닝 기본 알고리즘(RNN, CNN)과 이를 활용한 최신 알고리즘(Seq2seq, Attention)에 대해 함께 공유해서 챗봇 성능의 극대화를 위한 준비를 합니다.
- 참고자료 : [Word2Vec으로 문장 분류하기](#)
- 참고자료 : [CNN으로 문장 분류하기](#)
- 참고자료 : [딥러닝 기반 자연어처리 기법의 최근 연구 동향](#)
- 참고자료 : [딥러닝을 이용한 자연어 처리 입문](#)

■ 컨트리뷰톤 가이드

Step 4: 챗봇 프로젝트구축진행

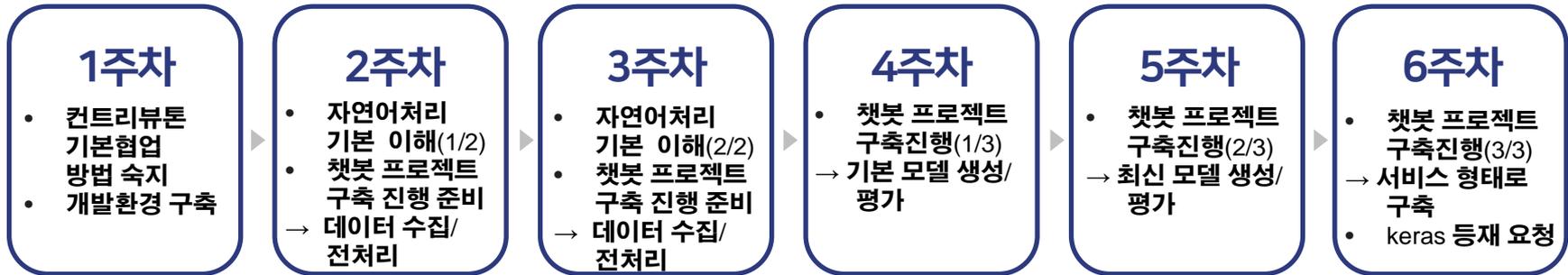
- 많은 사람들이 한글화된 챗봇을 쉽게 구현할 수 있도록 패키지 형태로 구축을 목표로 합니다.
- 데이터 수집, 전처리, 데이터 라벨링, 모델생성, 모델평가 등의 단계로 수행하나, 팀별의 역할은 2 단계에서 개인의 관심도나 역량에 따라 구분할 것입니다.
- 서비스, 모델, 데이터(DB)등이 유기적이고, 쉽게 구현할 수 baseline 작성을 목표로 합니다.
 - Babi 셋으로 케라스 공식 예제 활용 - 한글화 서비스 연동으로 실제 커뮤니티에 배포
 - 다양한 데이터 셋을 기반으로 일반/특정분야 문의로 구분하고 지속적인 형태로 구성
- 해당 카테고리 별로 labeling 작업 진행
- 적합한 모델 생성/ 평가 작업을 통해 모델 개발 수행
- 참고자료: [챗봇 구축 첫 미팅에서 자주 듣는 질문들](#)
- 참고자료 : [Creating a Chatbot from Scratch using Keras and TensorFlow](#)
- 참고자료 : [Seq2seq Chatbot for Keras](#)

■ 컨트리뷰톤 가이드

Step 5: 공식 Keras 페이지 등재 요청

- 협업해서 생성한 소스를 배포 합니다.
- 공식 Keras Repository에 등재하여 많은 사람들이 한글화된 우수한 챗봇을 쉽게 이용하고, 기여할 수 있도록 합니다.

■ 운영방안



온라인참여

- slack으로 이슈 위주의 커뮤니케이션 진행
- github에서 작업과정 기획 및 모니터링
- Zoom 화상회의 / 스크린 공유

오프라인 모임

- 선릉 개발자 센터 회의실 모임
- 필요에 따라 강남 부근에서 모임

규칙적인 상황 파악과 협업 진행을 위해 주 1회 온/오프라인 정기 모임 예정
개인 별 작업 진행 정도에 따른 보상(reward) 제공을 통해 멤버들의 사기/활력 유지

■ 멘토 소개



김태영

비전공자분들이 직관적이고 간결한 딥러닝 라이브러리인 케라스를 이용해 딥러닝 입문을 쉽게 할 수 있도록 <블록과 함께 하는 파이썬 딥러닝 케라스>의 집필과 '김태영의 케라스 블로그', '케라스 코리아', '캐글 코리아'를 운영하고 있습니다. 또한 '강화학습 코리아의 알파오목 프로젝트'의 팀원으로 알파고 모델을 대중들이 접할 수 있도록 서비스도 구축했습니다.
현재 (주)인스페이스에서 기술이사로서 "태양에서 세포까지 딥러닝", "게임에서 우주까지 강화학습"의 모토로 여러 분야 인공지능을 적용하고자 활발히 연구개발 하고 있습니다.



김영하

새로운 기술에 관심이 많은 엔지니어 성향을 가진 개발자 입니다. 전사 시스템 모니터링 및 빅데이터 플랫폼 기반의 실무 프로젝트를 수행했습니다. 프리랜서처럼 일하는 회사인 디플러스에서 데이터 분석, 인공지능 관련된 프로젝트 수행 및 강의를 하고 있습니다.
지난 2~7월간 진행된 MS Azure Discovery Day 행사인 2일차에 Azure 에서 하는 인공지능을 담당했습니다. 기술서를 읽고 공유하고자 공부하다 보니 어느덧 번역한 책이 '데이터 분석을 위한 판다스 입문', '파이썬 웹 스크래핑' 등 6권이 되었습니다. 재미있고 쉽게 누구나 해볼 수 있는 프로그래밍 및 인공지능 콘텐츠를 찾으며 만들고 있습니다



김슬기

두 분의 멘토 님의 책 읽는 것을 좋아하는 애독자 입니다. 머신러닝 엔지니어라는 잡 타이틀을 가지고, 이번 챗봇 프로젝트 멘토로 참여하게 되었습니다. 평소 케라스 코리아와 캐글 코리아에서 눈팅 하는 것을 즐기하며, 새로운 환경에서 새로운 것을 도전하는 것에 관심이 많습니다. 모두에게 의미있고, 즐거운 컨트리뷰톤이 되길 바랍니다.



감사합니다