

2021 Contribution Academy

Project Guide



NLP With U



01.

프로젝트 내용

프로젝트 분야 · 저장소 · 활용 언어 · 난이도
참가자 모집 유형



01. 프로젝트 내용

- 프로젝트명: NLP With U
- 프로젝트 분야: NLP, AI
- 프로젝트 저장소:
 - konlpy: <https://github.com/konlpy/konlpy>
 - kiwi: <https://github.com/bab2min/Kiwi>
 - 추후 활용 가능한 저장소
 - pororo: <https://github.com/kakaobrain/pororo>
 - Khaiii: <https://github.com/kakao/khaiii>
 - transformers: <https://github.com/huggingface/transformers>
 - opennmt: <https://github.com/OpenNMT/OpenNMT-py>
 - fairseq: <https://github.com/pytorch/fairseq>

01. 프로젝트 내용

- 활용 언어: Python
- 프로젝트 난이도: 중
- 참가자 모집 유형:
 - 기본적인 프로그래밍 능력이 있고, NLP에 관심이 있는 분
 - 오픈소스 활용 경험이 있고, 기여자가 되고 싶은 분
 - 스스로 잘하는 분보다는 도움이 필요한 분

01. 프로젝트 내용

- 우대사항:
 - Java, C++언어를 다룰 수 있는 분
 - konlpy, kiwi 등 NLP 도구를 사용해보고 경험을 정리해서 제출해주시는 분
 - NLP 도구를 실제로 사용 후 경험을 공유해주세요.
 - 예) <https://konlpy.org/ko/latest/#user-guide>
 - 위 내용은 자기소개서에 작성해주시면 됩니다.
 - 개발 블로그, github 링크가 있으면 공유해주세요.

02.

프로젝트 개요

프로젝트 상세 소개 내용



02. 프로젝트 개요

- NLP With U는 자연어처리에 쓰이는 다양한 도구들을 개선하고 발전시키기 위한 프로젝트입니다. 이미 영어권에서는 많은 수의 오픈소스 도구들이 공유되고 있지만, 한국어는 언어의 특성이 다르기에 이 도구들을 그대로 사용하기 어려운 경우가 많습니다. NLP With U는 이번 컨트리뷰션 아카데미를 통해 한국어 자연어처리를 위한 도구들에 작지만 의미 있는 기여를 하고자 합니다.
- konlpy는 대표적인 한국어 NLP 도구로, 형태소 분석과 품사 태깅을 비롯한 여러 기능을 제공하는 Python 패키지입니다.
- Kiwi는 konlpy에는 포함되지 않은 별개의 형태소 분석기로 다른 패키지에 의존성 없이 빠르고 효율적으로 동작하는 분석기를 목표로 합니다.
- 이외에도 다양한 자연어처리 도구가 있고 이들 모두 여러분들의 기여가 필요한 상황이기에, 컨트리뷰션 아카데미에서 기여할 저장소를 한정하지는 않았습니다. 논의를 통해 다른 저장소를 선정하여 함께 기여할 수 있습니다.

02. 프로젝트 개요

- 함께 기여해볼만한 저장소에는 다음과 같은 것들이 있습니다.
- 형태소 분석
 - khaiii: 카카오에서 공개한 신경망 기반의 형태소 분석기
- 통합
 - pororo: 카카오브레인에서 공개한 한국어를 비롯한 4개국어용 다양한 자연어처리 플랫폼
- 딥러닝 모델
 - Huggingface Transformers: Transformer 모델을 쉽게 사용할 수 있게 구성한 Python 패키지
 - OpenNMT: 기계 번역을 위한 Python 패키지
 - fairseq: Facebook에서 개발한 언어 모델링을 위한 Python 패키지
- 이외에도 오픈소스 자연어처리 도구라면 어떤 것이든 함께 기여해보고 싶어요~
- 머신러닝/딥러닝 등 복잡하고 어려운 모델이 아니라 단순한 도구도 좋습니다.
- 버그 수정 및 기능 추가뿐만 아니라 문서화, 오타수정, 이슈 처리와 같은 기여도 모두 좋아요!

03.

컨트리뷰션 가이드

단계별 컨트리뷰션 커리큘럼



03. 컨트리뷰션 가이드

- 코스 1. 컨트리뷰션에 필요한 Git/Github 사용법 학습
 - koNLPy, kiwi 등 한국어 자연어처리 도구를 사용하고 기여하기 위한 git/github 사용법을 학습합니다.
 - 기본적인 기능을 중심으로 학습하며 필요시에는 별도로 저장소를 생성하고 실습해봅니다.
 - 살펴볼 기능
 - fork, clone, merge, branch, commit, pr
 - 각 저장소마다 있는 contribution guide를 함께 살펴봅니다.

03. 컨트리뷰션 가이드

- 코스 2. 한국어 자연어처리 전반에 대한 간략한 소개
 - 한국어 자연어처리 기본 내용에 대해 소개합니다.
 - koNLPy, kiwi 등 한국어 자연어처리 도구를 사용을 위해 기본 개념에 대한 이해가 필요합니다.
 - 자연어처리란 무엇인가?
 - 자연어처리 분야 소개 (실제 업계의 적용 사례 위주)

03. 컨트리뷰션 가이드

- 코스 3. 다양한 한국어 자연어처리 도구 살펴보기 & 기여할 대상 선정하기
 - 한국어 자연어처리 도구를 함께 살펴보고, 기여할 대상을 선정합니다.
 - 꼭 한국어 자연어처리 도구가 아니더라도 자연어처리에 자주 사용되는 오픈소스를 모두가 기여 대상이 될 수 있습니다. 기여 대상은 다양한 경우가 있을 수 있습니다.
 - 예) 오타 수정, 이슈 처리, 기능 추가 등

03. 컨트리뷰션 가이드

- 코스 4. 개인별 기여 프로젝트 및 주제 논의, 프로젝트 구조 분석과 내부 이해
 - 개인별로 기여할 프로젝트를 정할 수 있도록 멘토링 합니다.
 - 주제를 못 찾을 경우 멘토가 다루고 있는 주제 중에서 간단한 주제들을 함께 해결을 제안합니다.
 - 기여할 대상에 대해 구조를 분석하고 내부를 함께 이해합니다.

03. 컨트리뷰션 가이드

- 코스 5. 지속적인 컨트리뷰션 독려 및 멘토링
 - 멘티가 진행중인 컨트리뷰션에 대해 지속적인 피드백을 줍니다.
 - 개인별로 처한 문제에 대해 함께 고민합니다.

03. 컨트리뷰션 가이드

- 코스 6. 컨트리뷰션 내용 정리 및 공유
 - 컨트리뷰션 아카데미를 통해 각자가 경험한 내용을 정리하고 공유하는 시간을 갖습니다.
 - 컨트리뷰톤이 잘 마무리 될 수 있도록 합니다.

04.

운영 방안

컨트리뷰션 단계별 상세 운영 방안



04. 운영 방안

| 1주차 | 2주차 | 3주차 | 4주차 | 5주차 |
|---------------------|----------------------|-----------------------|-------------------------|----------------|
| git & github에 대한 이해 | 한국어 자연어처리에 대한 간략한 소개 | 다양한 한국어 자연어처리 도구 살펴보기 | 멘티 관심사 파악 & 기여할 프로젝트 설정 | 프로젝트 구조 분석과 이해 |

[온라인]

- 코로나-19로 인해 기본적으로는 온라인 참여로 진행
- Google Meet, Slack, Github을 활용

[오프라인]

- 코로나 & 멘티 상황에 맞추어 필요에 따라 오프라인 모임 진행 (상세한 시간/장소는 조율 후 결정)

04. 운영 방안

| 6주차 | 7주차 | 8주차 | 9주차 | 10주차 |
|------------------------|------------------------|------------------------|------------------------|----------------------|
| 지속적인 컨트리뷰션 독려 및 멘토링 | 컨트리뷰션 내용 정리 및 공 유 |

[온라인]

- 기여 과정에서 겪는 어려움들을 코드 리뷰와 멘토링을 통해 온라인으로 풀어나갑니다.

[오프라인]

- 화상회의나 원격접속 등으로 충분치 않은 경우 언제든지 오프라인 미팅을 통해 문제를 해결하고 에너지를 얻어갑시다!

05.

멘토 소개

프로젝트팀 컨트리뷰션을 리딩 하실 멘토를 소개합니다!





이민철

카카오엔터프라이즈

- * 카카오엔터프라이즈 자연어처리팀 (2019~)
- * 한국어 형태소 분석기 Kiwi 개발 (2017~)



이다니엘

카카오엔터프라이즈

- * 오픈SW 파트 프런티어 (2020~)
- * 카카오/카카오엔터프라이즈 자연어처리팀 (2013~)

2021 Contribution Academy

