

2023 오픈소스 컨트리뷰션 아카데미

Open Source Contribution Academy

Hugging Face 한글화

 Project Guide


Open Source Contribution Academy 2023 Open Source Contribution Academy 2023 Open Source Contribution Academy 2023 Open Source Contribution Academy 2023 Open Source Contribution Academy



1

프로젝트 개요

프로젝트 분야 · 활용 언어 · Repository ·
난이도 · 참가자 모집유형 및 우대사항 등



1

프로젝트 개요

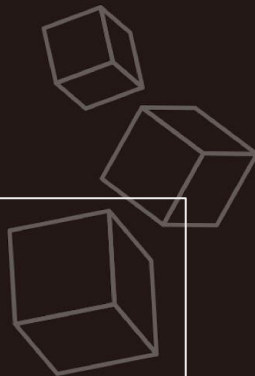
프로젝트명 : Hugging Face 한글화

프로젝트 분야 : AI, ML, NLP

프로젝트 저장소 : <https://github.com/huggingface>

활용 기술 : Python, Jupyter Notebook, Gradio, Markdown, Git, Shell Script

프로젝트 난이도 : 상중



*🤗 Hugging Face 팀과 직접 논의 중인 부분이 있어, 가이드의 최신 버전을 보려면 [이 링크](#)를 참조해주세요.

1

프로젝트 개요

참가자 모집 유형

- HuggingFace 오픈소스와 개발 생태계에 뛰어들고 싶은 분
- Transformers 기반 최신 ML 모델을 다뤄보고 싶은 분
- 기술 문서 등의 번역/요약/생성 태스크에 관심 있는 분
- 스스로 학습하고 배운 것을 공유하며 함께 성장하는 것에 익숙하신 분

우대 사항

- HuggingFace 또는 ML을 공부했거나 간단하게 응용해본 분
- 저녁, 주말 등의 시간을 공부하고 기여하는데 투자할 수 있는 분
- Python 코드 작성과 영어 대화에 거부감이 없으신 분들을 환영합니다.
- 기술 블로그를 운영한 경험이 있으신 분을 환영합니다.



2

프로젝트 소개

프로젝트 상세 소개 내용



2

프로젝트 소개

주요내용 1

허깅페이스 기술문서 한글화

허깅페이스의 공식 문서를 한글화 합니다.

허깅페이스에서 제공하는 기능 중 관심있는 분야의 문서를 한글로 번역합니다.

번역한 문서를 발표하는 과정을 통해서 허깅페이스에 대한 이해도를 높입니다.

주요내용 2

머신러닝 용어집

머신러닝은 빠르게 발전하는 분야로 다양한 머신러닝 용어가 만들어지고 있습니다.

한글화 작업 과정에서 추출된 용어에 대해서 이해하고 적절한 한국어 용어를 판단합니다.

머신러닝 용어집은 머신러닝 입문 개발자가 문서를 올바르게 이해할 수 있도록 도와줍니다.

2

프로젝트 소개

주요내용 3

허깅페이스 이해하기

허깅페이스 공식 문서를 발표하며 허깅페이스의 사용법과 기본 구조를 이해합니다. 발표과정와 함께 허깅페이스 글로벌 커뮤니티에 참여하여 자신이 관심있는 분야를 탐색합니다.

주요내용 4

허깅페이스 생태계에 기여하기

허깅페이스에서는 Model, Datasets, Space, Notebooks with examples과 같은 주제로 ML개발자들이 머신러닝 생태계에 기여할 수 있는 방법을 제공합니다. 관심 분야를 논의하고 주제를 선정하여 허깅페이스 생태계에 기여합니다.

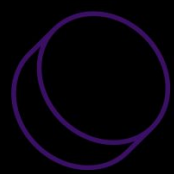


3

컨트리뷰션 가이드



단계별 컨트리뷰션 커리큘럼



3

컨트리뷰션 가이드

컨트리뷰션 코스 1

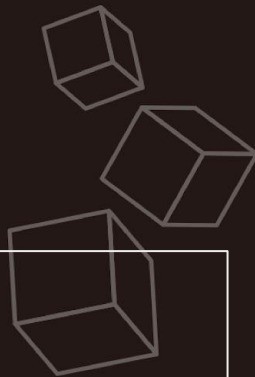
- 오픈소스 기여를 위한 GitHub 및 PR 생성 방법 익히기
 - 번역을 작성하고 반영하기 위한 여러 가지 Git, GitHub 사용법을 배웁니다.
 - 오픈소스의 기본인 Branch 관리 방법을 공부합니다.
 - PR을 서로 리뷰하고 수정하는 방법을 배웁니다.
- 기술문서 번역을 위한 커뮤니케이션 활동
 - Hugging Face 문서 배포에 사용되는 doc-builder를 사용해봅니다.
 - 문서 작성에 기본이 되는 Markdown 문법을 간단히 배웁니다.
 - 인공지능에 사용되는 중요한 용어를 한글로 어떻게 번역할지 논의하며 이해합니다.
 - 기술문서를 서로 리뷰하며 자연스럽게 읽기 편한 기술문서 작성법을 익힙니다.

3

컨트리뷰션 가이드

컨트리뷰션 코스 2

- Hugging Face에 대한 기본 개념 학습
 - Hugging Face 생태계의 기본 구조와 트랜스포머 모델 아키텍처에 대해 공부합니다.
 - 생태계의 라이브러리(Transformers, Datasets, Tokenizers, Accelerate)와 이를 사용한 자연어처리(NLP) 프로세스를 이해합니다.
 - 간단한 모델을 API로 다뤄보는 실습을 진행합니다.
- 머신러닝 용어집 이해 및 새로운 용어 도출
 - HuggingFace / Google ML Glossary로 중요 개념을 이해합니다.
 - 새롭게 맞닥뜨릴 머신러닝 용어의 한국어 번역에 진입장벽을 낮춥니다.



3

컨트리뷰션 가이드

컨트리뷰션 코스 3

- Hugging Face 라이브러리를 사용한 사이드 프로젝트 개발
 - 글로벌 커뮤니티에 참가하여 자신이 관심있는 프로젝트를 탐색합니다.
 - 한국 개발자의 요구사항에 맞는 Notebook Example로 문서화합니다.
 - 관심있는 인공지능 모델을 사용하여 블로그에 게시물을 작성합니다.
 - 🤗 Spaces에 직접 데모를 만들고, 커뮤니티에 공유합니다.
 - Gradio, Streamlit 등의 대시보드 툴을 활용하여 시연합니다.
 - 멘토-멘티, Hugging Face 글로벌 커뮤니티 피드백을 기반으로 프로젝트를 개선합니다.

* 🤗 Hugging Face 팀과 직접 논의 중인 부분이 있어, 가이드의 최신 버전을 보려면 [이 링크](#)를 참조해주세요.



4

컨트리뷰션 운영 방안

단계별 컨트리뷰션 커리큘럼

4 컨트리뷰션 운영 방안

1st Week

프로젝트 소개 (1 weeks)

1. 허깅페이스 소개
2. 프로젝트 및 Git Repository 생성
3. 커뮤니케이션 방식 소개
 - Notion, Slack, Discord 등 협업 툴 활용

2nd Week

프로젝트 소개 (2 weeks)

1. 개발 툴 스테디
 - Github, build-doc, transformers
 - Branch 관리 방법
2. 번역 Process 및 템플릿 소개
 - [\[18n-KO\] Translating docs to Korean](#)
3. 번역 개발환경 세팅 시작
 - [Generating the documentation](#)

4

컨트리뷰션 운영 방안

3rd Week

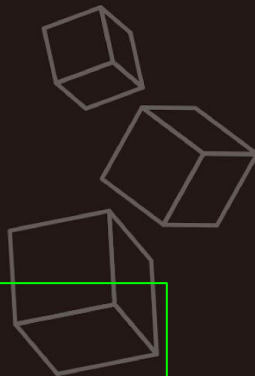
한글화 Sprint (1 weeks)

1. 번역 문서 및 발표문서 선정
 - 참고자료: [Hugging Face Transformers Docs](#)
2. Pull Request 프로세스 기반 컨트리뷰트
 - PR 가이드라인 제공
 - 참고자료: [Translating the Transformers documentation into your language](#)

4th Week

한글화 Sprint (2 weeks)

1. 허깅페이스 문서 번역 및 발표
2. 번역 Glossary 논의



4

컨트리뷰션 운영 방안

5th Week

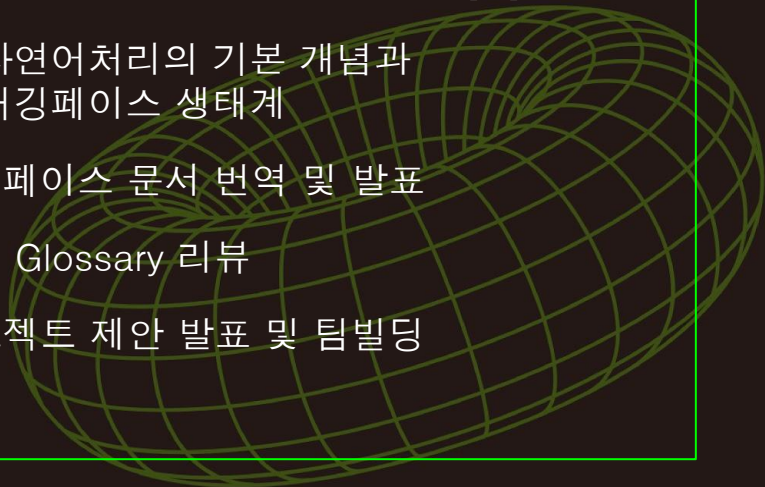
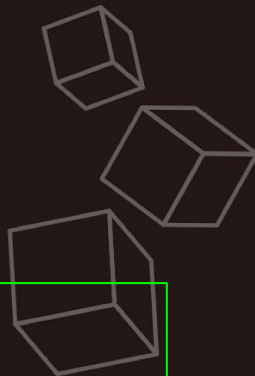
한글화 Sprint (3 weeks)

1. 허깅페이스 NLP course 스터디 [1]
 - 자연어처리의 기본 개념과 허깅페이스 생태계
2. 허깅페이스 문서 번역 및 발표
3. 번역 Glossary 리뷰
4. 프로젝트 탐색 및 정의

6th Week

한글화 Sprint (4 weeks)

1. 허깅페이스 NLP course 스터디 [1]
 - 자연어처리의 기본 개념과 허깅페이스 생태계
2. 허깅페이스 문서 번역 및 발표
3. 번역 Glossary 리뷰
4. 프로젝트 제안 발표 및 팀빌딩



4

컨트리뷰션 운영 방안

7th Week

허깅페이스 프로젝트 Sprint (1 weeks)

1. 허깅페이스 NLP course 스터디 [2]
 - Transformers 라이브러리 사용 방법
2. 프로젝트 진행 Sprint

8th Week

허깅페이스 프로젝트 Sprint (2 weeks)

1. 허깅페이스 NLP course 스터디 [2]
 - Transformers 라이브러리 사용 방법
2. 프로젝트 진행 Sprint
3. 번역 Glossary 기술블로그 결과 공유회

4

컨트리뷰션 운영 방안

9th Week

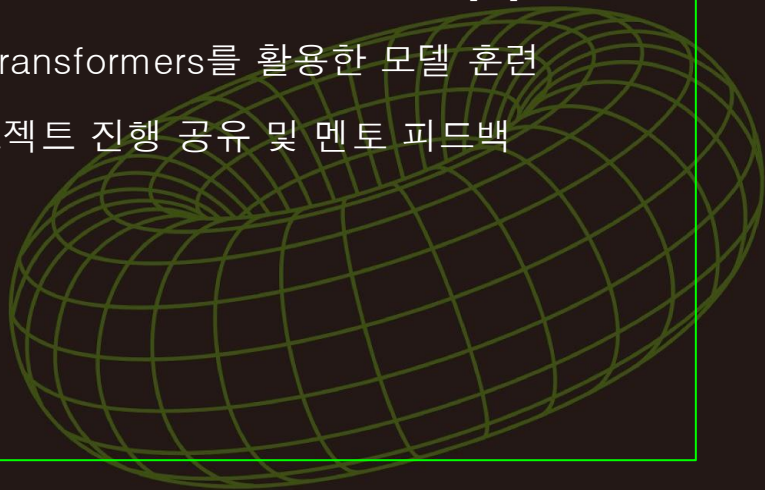
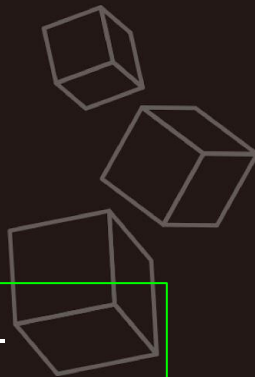
허깅페이스 프로젝트 Sprint (3 weeks)

1. 허깅페이스 NLP course 스터디 [3]
 - Transformers를 활용한 모델 훈련
2. 프로젝트 진행 공유 및 멘토 피드백

10th Week

허깅페이스 프로젝트 Sprint (4 weeks)

1. 허깅페이스 NLP course 스터디 [3]
 - Transformers를 활용한 모델 훈련
2. 프로젝트 진행 공유 및 멘토 피드백



4

컨트리뷰션 운영 방안

11th Week

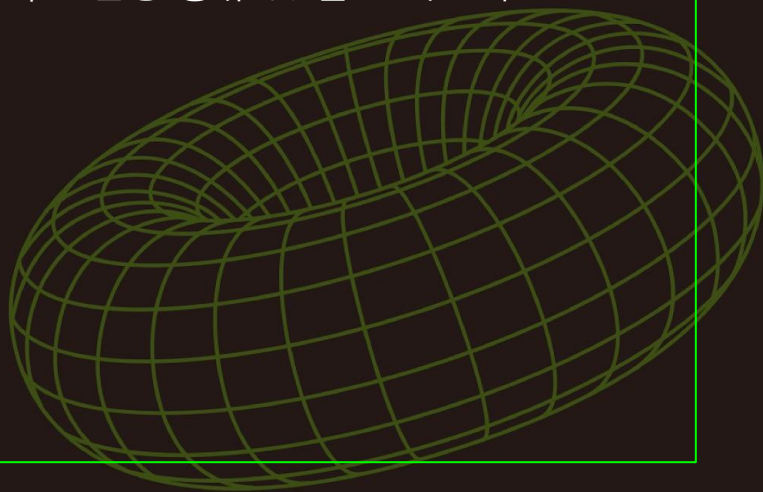
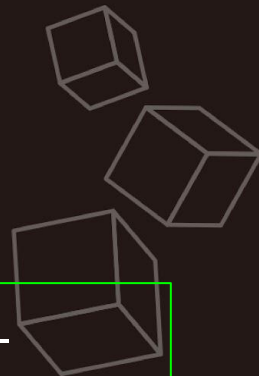
허깅페이스 프로젝트 Sprint (5 weeks)

1. 프로젝트 진행 공유 및 멘토 피드백
2. 프로젝트 중간 발표회

12th Week

허깅페이스 프로젝트 Sprint (6 weeks)

1. 프로젝트 진행 공유 및 멘토 피드백



4

컨트리뷰션 운영 방안

13th Week

허깅페이스 번역 및 프로젝트 회고 (1 week)

1. 기술블로그 Release

- 허깅페이스 번역 프로세스 문서
- 머신러닝 Glossary

2. 프로젝트 결과 공유

마무리

1. 프로젝트 결과 기술 블로그 작성
2. Kaggle 또는 Dacon 도전 계획 나누기
3. 허깅페이스 심화 스터디 논의하기
4. 머신러닝 Glossary 운영방안 논의

* 🤗 Hugging Face 팀과 직접 논의 중인 부분이 있어, 가이드의 최신 버전을 보려면 [이 링크](#)를 참조해주세요.

4

컨트리뷰션 운영 방안

◎ ONLINE

[번역 & 프로젝트]

- GitHub 기반 코드 리뷰, 피드백 및 버전 관리
- Google Meet를 사용하여 온라인 미팅 진행 (화면 공유 및 협업 가능)
- YouTube로 비공개 온라인 방송하여 원격 참여 및 다시보기 가능

[커뮤니티]

- 그룹 채팅 채널(카카오톡 팀채팅, 디스코드 등)을 통한 정기적인 소통과 업데이트 유지

◎ OFFLINE

[번역 & 프로젝트]

- OpenUP 서초 센터 또는 근처에서 사전 공지 후 오프라인 회의
- 프로젝트 발표회

[커뮤니티]

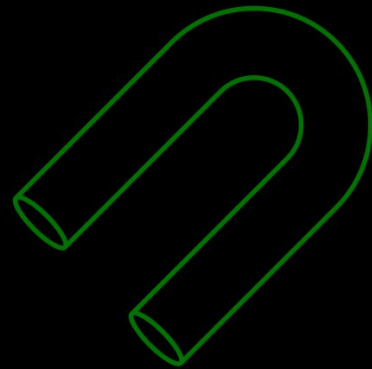
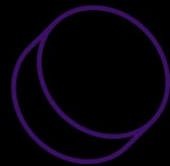
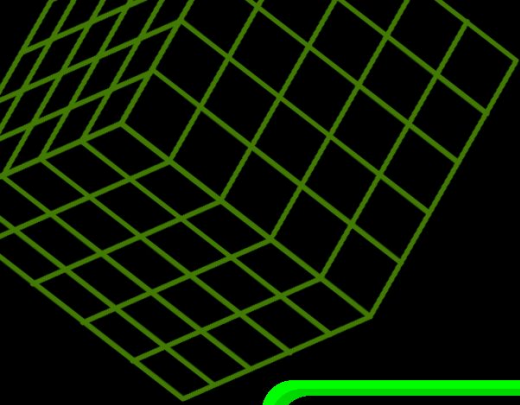
- 국내 오픈소스 커뮤니티 세미나 참여(온·오프라인)
- OpenUP 센터 또는 인근 지역에서 오프라인 미팅



5

멘토 소개

컨트리뷰션 프로젝트팀 멘토단 소개



5

멘토 소개



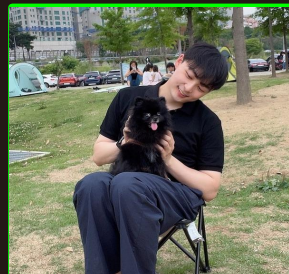
● 양성모

- 가짜연구소 6기 러너, 이달의 커뮤니케이터
- AI양재허브 인공지능 오픈소스 경진대회 1위 팀



● 윤현서

- 가짜연구소 6기 빌더
- Azurethon Cloud Skills Challenge #AI 2nd Prize



● 정우준

- 가짜연구소 6기 빌더



● 심소현

- 가짜연구소 5-6기 러너



● 서원형

- 가짜연구소 6기 러너
- Google ML 부트캠프 3기
- '22 컨트리뷰톤 AzureSDK 리드멘티

2023 오픈소스 컨트리뷰션 아카데미

Open Source Contribution Academy

Hugging Face 한글화

컨트리뷰션에 도전해 보세요!

 **THANK YOU** 

Open Source Contribution Academy 2023 Open Source Contribution Academy 2023 Open Source Contribution Academy 2023 Open Source Contribution Academy