Hugging Face Datasets

- Python_code_instructions_18k_alpaca
 - Instruction
 - Input
 - Output
 - Prompt
- Hugging Face Q Search models, datasets, users... Models Datasets ■ Datasets: @iamtarun/python code instructions 18k alpaca (□ | ♥ like 227 Tasks: Duestion Answering Start Text2Text Generation Text Generation Modalities: Text Formats: parquet Libraries:

 Datasets pandas Croissant +: **⊞** Dataset Viewer Auto-converted to Parquet (/> API Embed Full Screen Viewe train · 18.6k rows O Search this datase instruction string · lengths string · lengths string · lengths Create a function # Python code def sum sequence(sequence): sum to calculate the ... for num in sequence: sum += num return sum website: www.example.com data to crawl code for crawling_ crawl website for phone numbers(website): response Create a Pytho [x*x for x in [1, 2, 3, 5, 8, 13]] list comprehensio. Given a string, remove all the consecutive def remove duplicates(string): result = "" pre script to perform_ duplicates from the string. Input:_ ' for char in string: if char != prev: result Write a python def generate_random_divisible_number(): import Write a Python def third largest(lst): if len(lst) < 3: return [12, 13, 13, 45, 22, 99] code to get the ... distinct = [] for i in 1st: if i not in disting

- Column 소개

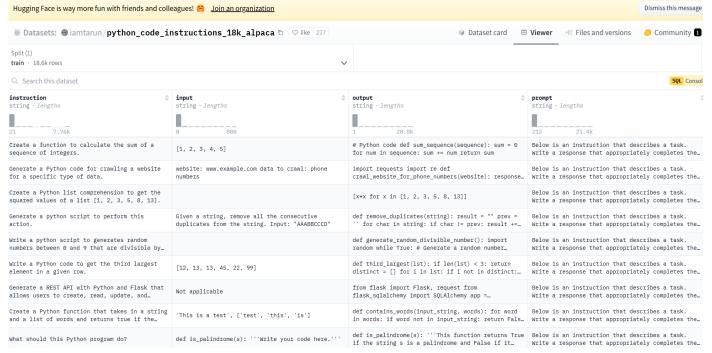
- Instruction (Task 지시사항)
- Input (Task 수행시 필요한 데이터)
- Output (Python 코드 출력)
- Prompt (질문 응답 텍스트)

- 예시

- Instruction = 시퀀스 합계를 계산하는 함수 만들어줘
- Input "[1, 2, 3, 4, 5, 6, ...]"
- Output "def_sum_sequence(sequence)..."
- Prompt

작업에 대한 요청이 있습니다.

요청을 완성하는 응답을 작성하십시오..





Previous 1 2 3 ... 187 Next >

Hugging Face Datasets

- Python_code_instructions_18k_alpaca
- 출처
 - Hugging Face Dataset
- 수집 방법
 - 프로젝트에 맞는 Code 생성을 위해 Hub에서 가장 최적화된 데이터셋에서 채택함

- 데이터 출처에 대한 법적 규제

- 하단의 이미지는 Hugging Face Docs Datasets의 내용 입니다.
- 데이터 수집 목적
 - 모델의 학습을 위해서 데이터 수집을 해야 했고 방법에 기술 되어 있듯이 Code 생성을 위해 (메인인 Python을 위해) 최대한 잘 정제되어있는 데이터셋을 가져옴

Datasets



Datasets is a library for easily accessing and sharing datasets for Audio, Computer Vision, and Natural Language Processing (NLP) tasks.

Load a dataset in a single line of code, and use our powerful data processing methods to quickly get your dataset ready for training in a deep learning model. Backed by the Apache Arrow format, process large datasets with zero-copy reads without any memory constraints for optimal speed and efficiency. We also feature a deep integration with the Hugging Face Hub, allowing you to easily load and share a dataset with the wider machine learning community.

Find your dataset today on the <u>Hugging Face Hub</u>, and take an in-depth look inside of it with the live viewer.

Tutorials

Learn the basics and become familiar with loading, accessing, and processing a dataset. Start here if you are using Datasets for the first time!

How-to guides

Practical guides to help you achieve a specific goal. Take a look at these guides to learn how to use Datasets to solve real-world problems.

Conceptual guides

High-level explanations for building a better understanding about important topics such as the underlying data format, the cache, and how datasets are generated.

Reference

Technical descriptions of how

Datasets classes and methods work.

