



What if my dataset isn't on the Hub?

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive



What if my dataset isn't on the Hub?

Install the Transformers, Datasets, and Evaluate libraries to run this notebook.

```
[ ] !pip install datasets evaluate transformers[sentencepiece]
```

```
[ ] !wget https://github.com/crux82/squad-it/raw/master/SQuAD_it-train.json.gz
!wget https://github.com/crux82/squad-it/raw/master/SQuAD_it-test.json.gz
```

```
[ ] !gzip -dkv SQuAD_it-*.json.gz
```

```
[ ] from datasets import load_dataset

squad_it_dataset = load_dataset("json", data_files="SQuAD_it-train.json", field="data")
```

```
[ ] squad_it_dataset
```

```
DatasetDict({
  train: Dataset({
    features: ['title', 'paragraphs'],
    num_rows: 442
  })
})
```

```
[ ] squad_it_dataset["train"][0]
```

```
{
  "title": "Terremoto del Sichuan del 2008",
  "paragraphs": [
    {
      "context": "Il terremoto del Sichuan del 2008 o il terremoto...",
      "qas": [
        {
          "answers": [{"answer_start": 29, "text": "2008"}],
          "id": "56cdca7862d2951400fa6826",
          "question": "In quale anno si è verificato il terremoto nel Sichuan?",
        },
        ...
      ],
    },
    ...
  ],
}
```

```
[ ] data_files = {"train": "SQUAD_it-train.json", "test": "SQUAD_it-test.json"}
squad_it_dataset = load_dataset("json", data_files=data_files, field="data")
squad_it_dataset
```

```
DatasetDict({
  train: Dataset({
    features: ['title', 'paragraphs'],
    num_rows: 442
  })
  test: Dataset({
    features: ['title', 'paragraphs'],
    num_rows: 48
  })
})
```

```
[ ] data_files = {"train": "SQUAD_it-train.json.gz", "test": "SQUAD_it-test.json.gz"}
squad_it_dataset = load_dataset("json", data_files=data_files, field="data")
```

```
[ ] url = "https://github.com/crux82/squad-it/raw/master/"
data_files = {
  "train": url + "SQUAD_it-train.json.gz",
```

```
        "test": url + "SQuAD_it-test.json.gz",  
    }  
    squad_it_dataset = load_dataset("json", data_files=data_files, field="data")
```