



ELIXIR DATASETS

HUB OF READY-TO-USE DATASETS FOR ML MODELS

WERONIKA WOJTAS

RADOSŁAW ROLKA

PROMOTOR: DR INŻ. ALEKSANDER SMYWIŃSKI-POHL

KONCEPCJA ROZWIAZANIA

Stworzenie intuicyjnej w użyciu biblioteki w języku Elixir,
która umożliwia pobieranie oraz zarządzanie zbiorami
danych wykorzystywanymi w projektach związanych ze
sztuczną inteligencją.

WZOROWANE ROZWIAZANIE

Hugging face - Datasets



Ale w ekosystemie



WYMAGANIA

- Możliwość pobierania datasetów z zewnętrznych źródeł
- Zapisywanie datasetów na komputerze
- Wydajny system cashowania i ładowania zbiorów danych
- Pełna dokumentacja funkcji biblioteki

PRZYKŁAD DOKUMENTACJI W EXDOCU

```
@doc """
Loads a dataset from the given repository.

The repository can be either a local directory or a Hugging Face repository.

## Options

* `{:auth_token}` - the token to use as HTTP bearer authorization
  for remote files. If not provided, the token from the
  `ELIXIR_DATASETS_HF_TOKEN` environment variable is used.

## Returns

An `{:ok, %{dataset: paths}}` tuple, where `paths` is a list of
paths to the downloaded dataset files. If the dataset cannot be
loaded, an `{:error, reason}` tuple is returned.
If the dataset is not found, an error is raised.

## Examples

todo
"""

@spec load_dataset(repository(), keyword()) :: 
    {:ok, %{dataset: [Path.t()]}} | {:error, String.t()}
def load_dataset(repository, opts \\ []) do
  repository = normalize_repository!(repository)

  with {:ok, repo_files} <- get_repo_files(repository),
       {:ok, paths} <- maybe_load_model_spec(opts, repository, repo_files) do
    {:ok, %{dataset: paths}}
  end
end
```

Installation

If available in Hex, the package can be installed by adding `elixir_datasets` to your list of dependencies in `mix.exs` :

```
def deps do
  [
    {:elixir_datasets, "~> 0.1.0"}
  ]
end
```

Documentation can be generated with ExDoc and published on [HexDocs](#). Once published, the docs can be found at https://hexdocs.pm/elixir_datasets.

INSTRUKCJA INSTALACJI

PRZYKŁAD UŻYCIA W LIVEBOOKU

Setup ▾

```
1 ##### Target version
2 # Install dependencies
3 # Mix.install([
4 #   {:elixir_datasets, git: "https://github.com/radoslawrolka/ElixirDatasets"}
5 # ])
6 ##### Local dev-testing version
7 Mix.install([
8   {:elixir_datasets, path: "#{__DIR__}/.."},
9   {:explorer, "0.10.0"}
10 ],
11 config: Config.Reader.read!("#{__DIR__}/../config/config.exs")
12 )
13
14 # get auth_token for downloading from HuggingFace
15 auth_token = Application.get_env(:elixir_datasets, :hf_token)
16 :ok
```

+ Python

Check import

```
1 import ElixirDatasets
2
3 {:ok, dataset} = ElixirDatasets.load_dataset(
4   {:hf, "fka/awesome-chatgpt-prompts"},%
5   %{auth_token: auth_token})
```

PRZEBIEG PRAC

commit timeline

Faza Koncepcyjna i Planowania Architektury



Faza Implementacji



Faza Testowania, Optymalizacji i Dokumentacji



ARCHITEKTURA

zdjęcie diagramu uml

TESTOWANIE

Market Growth Rate

image

image

19.5%

Github Actions

Mix Test

METODYKA PRACY

- Spotkania z klientem
- Code Review
- CI/CD
- Testowanie
- Refactoring
- TDD

image

FILMIK JAK DZIAŁA TO W LIVEBOOKU

ZESPÓŁ



Weronika Wojtas



Radosław Rolka