

Package ‘hfhub’

August 18, 2023

Title Hugging Face Hub Interface

Version 0.1.1

Description Provides functionality to download and cache files from 'Hugging Face Hub' <<https://huggingface.co/models>>.

Uses the same caching structure so files can be shared between different client libraries.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports httr, filelock, fs, cli, withr, curl, glue, urltools, rlang

Suggests testthat (>= 3.0.0), jsonlite

Config/testthat/edition 3

URL <https://mlverse.github.io/hfhub/>

NeedsCompilation no

Author Daniel Falbel [aut, cre],
Posit [cph]

Maintainer Daniel Falbel <daniel@posit.co>

Repository CRAN

Date/Publication 2023-08-18 19:22:33 UTC

R topics documented:

hub_download	2
hub_repo_info	3
hub_snapshot	3
WEIGHTS_NAME	4
Index	6

hub_download	<i>Downloads files from HuggingFace repositories</i>
--------------	--

Description

Downloads files from HuggingFace repositories

Usage

```
hub_download(
  repo_id,
  filename,
  ...,
  revision = "main",
  repo_type = "model",
  local_files_only = FALSE,
  force_download = FALSE
)
```

Arguments

repo_id	The repository identifier, eg "bert-base-uncased" or "deepset/sentence_bert".
filename	Filename to download from the repository. Example "config.json".
...	currently unused.
revision	Revision (branch, tag or commitid) to download the file from.
repo_type	The type of the repository. Currently only "model" is supported.
local_files_only	Only use cached files?
force_download	For re-downloading of files that are cached.

Value

The file path of the downloaded or cached file. The snapshot path is returned as an attribute.

Examples

```
try({
  withr::with_envvar(c(HUGGINGFACE_HUB_CACHE = tempdir()), {
    path <- hub_download("gpt2", "config.json")
    print(path)
    str(jsonlite::fromJSON(path))
  })
})
```

hub_repo_info	<i>Queries information about Hub repositories</i>
---------------	---

Description

Queries information about Hub repositories

Usage

```
hub_repo_info(  
  repo_id,  
  ...,  
  repo_type = NULL,  
  revision = NULL,  
  files_metadata = FALSE  
)
```

```
hub_dataset_info(repo_id, ..., revision = NULL, files_metadata = FALSE)
```

Arguments

repo_id	The repository identifier, eg "bert-base-uncased" or "deepset/sentence_bert".
...	currently unused.
repo_type	The type of the repository. Currently only "model" is supported.
revision	Revision (branch, tag or commitid) to download the file from.
files_metadata	Obtain files metadata information when querying repository information.

Functions

- `hub_dataset_info()`: Query information from a Hub Dataset

hub_snapshot	<i>Snapshot the entire repository</i>
--------------	---------------------------------------

Description

Downloads and stores all files from a Hugging Face Hub repository.

Usage

```

hub_snapshot(
  repo_id,
  ...,
  revision = "main",
  repo_type = "model",
  local_files_only = FALSE,
  force_download = FALSE,
  allow_patterns = NULL,
  ignore_patterns = NULL
)

```

Arguments

<code>repo_id</code>	The repository identifier, eg "bert-base-uncased" or "deepset/sentence_bert".
<code>...</code>	currently unused.
<code>revision</code>	Revision (branch, tag or commitid) to download the file from.
<code>repo_type</code>	The type of the repository. Currently only "model" is supported.
<code>local_files_only</code>	Only use cached files?
<code>force_download</code>	For re-downloading of files that are cached.
<code>allow_patterns</code>	A character vector containing patterns that are used to filter allowed files to snapshot.
<code>ignore_patterns</code>	A character vector containing patterns to reject files from being downloaded.

<code>WEIGHTS_NAME</code>	<i>Weight file names in HUB</i>
---------------------------	---------------------------------

Description

Weight file names in HUB

Usage

```
WEIGHTS_NAME()
```

```
WEIGHTS_INDEX_NAME()
```

Value

A string with the default file names for indexes in the Hugging Face Hub.

Functions

- `WEIGHTS_NAME()`: Name of weights file
- `WEIGHTS_INDEX_NAME()`: Name of weights index file

Examples

```
WEIGHTS_NAME()  
WEIGHTS_INDEX_NAME()
```

Index

hub_dataset_info (hub_repo_info), 3
hub_download, 2
hub_repo_info, 3
hub_snapshot, 3

WEIGHTS_INDEX_NAME (WEIGHTS_NAME), 4
WEIGHTS_NAME, 4