

# Tsinghua Cloud Downloader

清华云盘批量下载助手，适用于分享的文件 size 过大导致无法直接下载的情况，本脚本添加了更多实用的小功能：

- ☒ 直接下载链接中的所有文件，无打包过程，且无文件数量和大小限制
- ☒ 支持下载带密码云盘链接
- ☒ 支持查看文件下载总大小和下载进度
- ☒ 支持模糊匹配需要下载的文件（如指定文件类型 / 指定文件夹下载）
- ☒ 支持单个文件（f）链接和文件夹（d）链接
- ☒ 支持自定义保存路径和自定义下载文件/文件夹名称

## Dependency

需要提前安装 python，安装过程略，以及 requirements.txt 文件里面的依赖库：

```
pip install -r requirements.txt
```

## Usage

Flags	Default	Description
<code>--link, -l</code>	<b>Required</b>	Share link of Tsinghua Cloud.
<code>--save_dir, -s</code>	<code>~/Desktop</code>	Path to save the files. Default: Desktop
<code>--file, -f</code>	None	Regex to match the file path. Default: download all files.
<code>--password, -p</code>	None	Password for the share link, if needed.
<code>--name, -n</code>	None	Custom name for the downloaded single file (for single-file share link) or the root directory (for directory share link). If not provided, the original name will be used.

## Example

```
python thu_cloud_download.py \
  -l https://cloud.tsinghua.edu.cn/d/1234567890/ \
  -s "/PATH/TO/SAVE" \
  -f "*.pptx?" (regex, 正则表达式) \
  -p "password" (if needed) \
  -n "custom_dir_name" (if needed)
```

## Support file format

`--file, -f` 参数支持 UNIX shell 风格的 pattern 字符串，支持使用如下几个通配符：

- `*`: 可匹配任意个任意字符。
- `?`: 可匹配一个任意字符。
- `[字符序列]`: 可匹配中括号里字符序列中的任意字符。该字符序列也支持中画线表示法。比如 `[a-c]` 可代表 a, b 和 c 字符中任意一个。
- `[!字符序列]`: 可匹配不在中括号里字符序列中的任意字符。

具体用法如下：

```
# 下载链接中所有文件
python thu_cloud_download.py -l https://xxx
# 下载链接中所有的.txt文件
python thu_cloud_download.py -l https://xxx -f *.txt
# 下载链接中某个文件夹的所有文件
python thu_cloud_download.py -l https://xxx -f folder/subfolder/*
```

## Output Log Example

下载链接中的全部 `txt` 文件：

```
>> python thu_cloud_download.py -l
https://cloud.tsinghua.edu.cn/d/b9aca92417f04166acdc/ -f *.pcd

2023-12-18 21:15:11,853 - INFO - Share key: b9aca92417f04166acdc
2023-12-18 21:15:12,811 - INFO - Searching for files to be downloaded, Wait
a moment...
=====
Last Modified Time           File Size   File Path
-----
2022-06-27T19:30:21+08:00    53324648   /cam3.pcd
2022-06-27T19:30:28+08:00    52693930   /cam4.pcd
2022-06-27T19:30:35+08:00    52672991   /cam5.pcd
2022-06-27T19:30:42+08:00    52774114   /cam6.pcd
-----
2023-12-18 21:15:14,449 - INFO - # Files: 4. Total size: 201.7 MB.
Start downloading? [y/n]y
2023-12-18 21:15:25,671 - INFO - Root directory name: livoxscan_20220626
[4/4]: 100%|████████████████████████████████████████| 202M/202M [00:33<00:00,
6.26MiB/s]
2023-12-18 21:15:59,479 - INFO - Download finished.
```