Twitter下载地址：

<https://figshare.com/articles/dataset/PHEME_rumour_scheme_dataset_journalism_use_case/2068650/2>

使用说明：

1. **文件夹Threads->en中有8个谣言数据：**

Charliehebdo, ebola-essien, ferguson, germanwings-crash, ottawashooting, prince-toronto, putinmissing, sydneysiege。

1. **每个话题中包含若干个子话题，每个子话题中包含文件**：

（1）image（存储该对话树中所有推文（包括源推文和所有回复）所包含的图片文件。）

（2）reactions（包含了所有对“源推文”的回复推文 (reply tweets) 或 引用推文 (quote tweets) 的完整JSON数据。文件夹里的每一个.json文件都对应一条回复推文。）

（3）source-tweets（包含引发整个对话树的那条原始推文的完整JSON数据）

（4）url-content（存储了推文中出现的外部链接 (URLs) 所指向的网页内容。）

（5）annotation.json（人工标注文件，也是整个数据集中最重要的文件之一。它以JSON格式记录了对源推文和部分回复推文的“真相”标签。）

（6）images.dat（一个映射文件，记录了哪条推文ID包含了哪个图片文件。）

（7）retweets.json（记录了谁转发了源推文。它通常包含一个转发用户ID的列表。）

（8）structure.json（描述了整个对话树的结构。它以JSON格式定义了推文之间的父子关系（即哪条推文回复了哪条推文））

（9）who-follows-whom.dat（记录了参与这个对话树的用户之间的关注关系。）

**3.在每个JSON文件中，需要重点关注以下字段信息：**

（1） 关于推文本身

"text"（该推文的文本信息），

"in\_reply\_to\_status\_id"（该推文转发的原推文id）,

"id"（该推文的id），

"created\_at"（推文的创建时间），

（2）推文的互动与状态

"retweet\_count"（这条推文被转发的次数）

"favorite\_count"（这条推文被“喜欢”的次数。）

（3）关于发推用户："user"对象包含了发布这条推文的用户的详细信息。其中：

"id"（用户的唯一ID），

"followers\_count"（粉丝数）

"friends\_count"（关注数（他关注的人），

"listed\_count"（该用户被添加到公开列表的次数）

"statuses\_count"（该用户发布的推文总数（包括转推）

—————————————————每个JSON中字段的意思————————————