

973 数据说明 Jasmine Revolution

数据下载来源为 Twitter, Flickr, YouTube, Google news 和 Qwiki 下面分别对下载的数据说明。

数据库名: jasminerevolution

1. Twitter

1) 数据库表内容属性

数据库表名: JasmineRevolution twitterdata

属性主要是以下信息:

number: 主键属性国家_事件号_条目号

content: tweet 内容

created_at: 创建时间

hits: 被点击次数

author: 作者

url: 对应的 url

2) 下载数据统计

国家	事件数	记录条数
Egypt	12	277
Jordan	9	79
Libya	13	509
saudi arabia	7	215
Syria	13	387
Tunisia	10	106
Yemen	12	544

3) 数据库使用说明

```
CREATE TABLE twitterdata(number VARCHAR(100) NOT NULL  
PRIMARY KEY, content TEXT, created_at VARCHAR(30), hits INT ,author  
VARCHAR(100), url VARCHAR(200));
```

mysqldump 在安装文件的 bin 目录下

mysqldump -u 用户名 -p 数据库名>导出的文件名

mysqldump -u root -p jasminerevolution>twitterdata.sql

mysql>use 数据库 (建立和导入数据库同名的数据库)

mysql>source twitterdata.sql(导出的 sql 文件所在的目录)

2. Flickr

1) 数据库表内容属性

数据库表名: JasmineRevolution flickrdata

属性主要是以下信息:

数据记录属性格式:

num_label(国家_事件号_条目号) image_id(图片 id) title(标题)
tags (tag) description(描述) time_taken(发生时间) time_upload (上传时间) latitude(纬度) longitude (经度) url_m url_o userid (用户)
url_s (源地址)

2) 下载数据统计

总计 68 事件 42704 图片

国家	事件数	记录条数
Egypt	18	9684
Jordan	6	1207
Libya	6	11129
saudi arabia	7	420
Syria	16	11912
Tunisia	4	4571
Yemen	11	3781

3) 数据库使用说明

```
CREATE TABLE flickrdata (num_label INT NOT NULL PRIMARY KEY ,  
image_id VARCHAR(20) NOT NULL, title TEXT, tags TEXT, description  
TEXT, time_taken VARCHAR(25), time_upload VARCHAR(25), latitude  
VARCHAR(15),longitude VARCHAR(15), url_m VARCHAR(100),url_o  
VARCHAR(100),userid VARCHAR(20),url_s VARCHAR(100) );
```

mysqldump 在安装文件的 bin 目录下

mysqldump -u 用户名 -p 数据库名>导出的文件名(有可能需要以管理员身份运行 cmd.exe)

```
mysqldump -u root -p jasminerevolution>flickrdata.sql
```

```
mysql>use 数据库 (建立和导入数据库同名的数据库)
```

```
mysql>source flickrdata.sql(导出的 sql 文件所在的目录)
```

3. Youtube

1) 数据库表内容属性

数据表名: **Jasminerevolution Youtubedata**

数据说明及数据库建立

(1) URLs.txt 文件包含 video 的 6 个属性: video 的 ID, video 的 URL, 事件号, 搜索关键词, video 的 title, video 的 tag, 他们在数据库中的名称分别为:

videoid, URL, country_event_num, serchKeyword, videoTitle, videoKeyword

(2) 视频的命名规则为 URL 的一部分, 从"...watch?v="之后的字符到结束或者"&..."之前。如:

对于 URL "http://www.youtube.com/watch?v=gMbl38ni0I4", 其视频命名为"gMbl38ni0I4"

对于 URL

"http://www.youtube.com/watch?v=a2_m90Ug_rA&feature=results_main&playlistnext=1&list=PL81F841C9D8235228", 其视频命名为"a2_m90Ug_rA"

2) 下载数据统计

总计 **434 videos**

国家	事件数	记录条数
Egypt	18	75
Jordan	10	48
Libya	18	69
saudi arabia	11	41
Syria	16	56
Tunisia	17	92
Yemen	16	53

3) 数据库使用说明

数据表建立

```
CREATE TABLE YouTubedata( videoid INT,URL VARCHAR(100),  
county_event_num VARCHAR(30),searchKeyword VARCHAR(100),  
videoTitle VARCHAR(200),videoTags TEXT(150));
```

数据表导入数据库

```
LOAD DATA LOCAL INFILE 'd:\URLs.txt' INTO TABLE YouTubedata  
FIELDS TERMINATED BY '|';
```

4. Google News

1) 数据库表内容属性

数据库表名: JasmineRevolution GoogleNews

属性主要是以下信息:

number: 主键属性国家_事件号_条目号

title: 新闻标题

publisher: News 的发布者, 一般表示某网站、该媒体机构

publisherDate: 发布日期

content: 内容

articleLanguage: 文档语言

articleURL: url

imageName: 图像名字

imageURL: 图像 url

2) 下载数据统计

共计 686 个新闻数

国家	事件数	记录条数
Egypt	17	144
Jordan	16	67
Libya	20	106
saudi arabia	12	71
Syria	15	88
Tunisia	31	104
Yemen	21	115

3) 数据库使用说明

```
CREATE TABLE GoogleNews(number VARCHAR(50) NOT NULL PRIMARY KEY, title VARCHAR(500) NOT NULL, publisher VARCHAR(500) NOT NULL, publishedDate VARCHAR(500) NOT NULL, content TEXT NOT NULL, articlelanguage VARCHAR(100) NOT NULL, articleURL VARCHAR(500) NOT NULL, imageName VARCHAR(500) NOT NULL, imageURL VARCHAR(500) NOT NULL);
```

mysqldump 在安装文件的 bin 目录下

mysqldump -u 用户名 -p 数据库名>导出的文件名

mysqldump -u root -p jasminerevolution>googlenews.sql

mysql>use 数据库 (建立和导入数据库同名的数据库)

mysql>source googlenews.sql (导出的 sql 文件所在的目录)

5. Qwiki

Qwiki 是一个创造交互性，即时生成，展现多媒体信息的平台。关于 JasmineRevolution, Qwiki 的数据虽有一定结构化但呈现的信息少。每个国家包含革命事件的 scrip、代表性的图片、和合成音频。

下载数据统计

国家	图片数	音频数
Egypt	27	9
Jordan	7	4
Libya	18	7
saudi arabia	15	7
Syria	22	10
Tunisia	32	5
Yemen	20	10