

1 Introduction

The “textit-downloader” is a server developed in Java, in order to retrieve the data from [TextIt](#) service every week, using [TextIt API](#).

See WSCTextItGuidev1.pdf (by Tyler) for more details about TextIt.

2 Implementation

2.1 Data retrieval

The server gets the flow ids from TextIt first, then downloads the runs according to the flow ids.

2.1.1 Getting TextIt flow id

It gets flow ids from TextIt using [TextIt API Flow Endpoint](#): <https://api.textit.in/api/v1/flows.json>

As to the requirement of weekly retrieval, it only collects the data of the last two weeks by adding the query parameter “after”: <https://api.textit.in/api/v1/flows.json?after=2016-01-18T00:00:00.000>

Then the TextIt service would respond a list of the flows in JSON format.

The server searches the “uuid” of the flows in the response and saves them in a list.

2.1.2 Downloading data of the runs

According to the list, the server downloads the runs by using [TextIt API Flow Run Endpoint](#): https://api.textit.in/api/v1/runs.json?flow_uuid=f5901b62-ba76-4003-9c62-72fdacc1b7b7 , and then save them as JSON files named as follow,

20151026-0c07e01a-3123-4359-9436-d43e4fb44043-1-runs.json

20151026 : The JSON file was downloaded on Oct. 26, 2015.

0c07e01a-3123-4359-9436-d43e4fb44043 : The UUID of the flow. (flow_uuid) I think this is the unique id for flows.

1 : This file is the first page of the runs data.

2.1.3 Downloading data of the flows

According to the list, the server downloads the runs by using [TextIt API Flow Endpoint](#): <https://textit.in/api/v1/flows.json?after=2016-05-10T00:00:00.000&&before=2016-05-16T23:59:59.000> and then save them as JSON files named as follows.

20160510-20160516-1-flows.json

This JSON is Weekly download sample run i.e from 2016-05-10 to 2016-05-16.

1 : This file is the first page of the flow data.

2.1.4 Downloading data of the contacts

According to the list, the server downloads the runs by using [TextIt API Contacts Endpoint](#): <https://textit.in/api/v1/contacts.json?after=2016-05-10T00:00:00.000&&before=2016-05-16T23:59:59.000> , and then save them as JSON files named as follow,

20160510-20160516-2-contacts.json

This JSON is Weekly download sample run i.e from 2016-05-10 to 2016-05-16

2 : This file is the second page of the contacts data.

See “public class TextItDownloader” in TextItDownloader.java for more details. (Line 27 – 29)

2.2 Loading data to MongoDB :

In MongoDB we have created WSC_RAW database. We store data of 3 endpoints i.e. flows, runs and contacts as 3 separate collections.

For instance, below are the sample collections present in WSC_RAW database.

Kenya_daily_20160516contacts

Kenya_daily_20160516flows

Zambia_daily_20160516contacts

Zambia_daily_20160516flows

Zambia_daily_20160516runs

Zambia_weekly_20160516contacts

Zambia_weekly_20160516flows

Zambia_weekly_20160516runs

2.3.1 Daily & Weekly running

This server runs Daily retrieve the data of the TextIt flows that created in the current day by setting up the job scheduler *cron* in Linux system as follows

```
00 08 * * * textit-downloader/WSC/bin/start-downloader.sh /u/smallag/textit-downloader/WSC/conf/global_zambia.properties
00 02 * * * textit-downloader/WSC/bin/start-downloader.sh /u/smallag/textit-downloader/WSC/conf/global_kenya.properties
00 09 * * * textit-downloader/WSC/bin/move_to_mongodb.sh /u/smallag/textit-downloader/WSC/conf/global_zambia.properties
00 03 * * * textit-downloader/WSC/bin/move_to_mongodb.sh /u/smallag/textit-downloader/WSC/conf/global_kenya.properties
```

This server also runs weekly, retrieve the data of the TextIt flows that created in the last week.

For example, to get the weekly data of Zambia’s flows, the server runs on Monday’s at 2 pm (Zambia time).

For Kenya’s weekly data retrieval is on Saturday at 9 am (Kenya time).

2.4 Deploying

After running “mvn install”, find the “target/deploy” folder in the project folder and upload the “bin”, “conf”, and “lib” folders and the jar file in it to the VM.

Fill up the value of token in “conf/global_zambia.properties” and “conf/global_kenya.properties” file.

If you would like to store the JSON files in a specific place, change the value of outputdir in “conf/global_zambia.properties” and “conf/global_kenya.properties” file.

For MongoDB executions , Fill up the value of credentials in “conf/global_zambia.properties” and “conf/global_kenya.properties” file.

Set up the job scheduler “cron” as mentioned above.