

Discussion board Notification Tool Tutorial

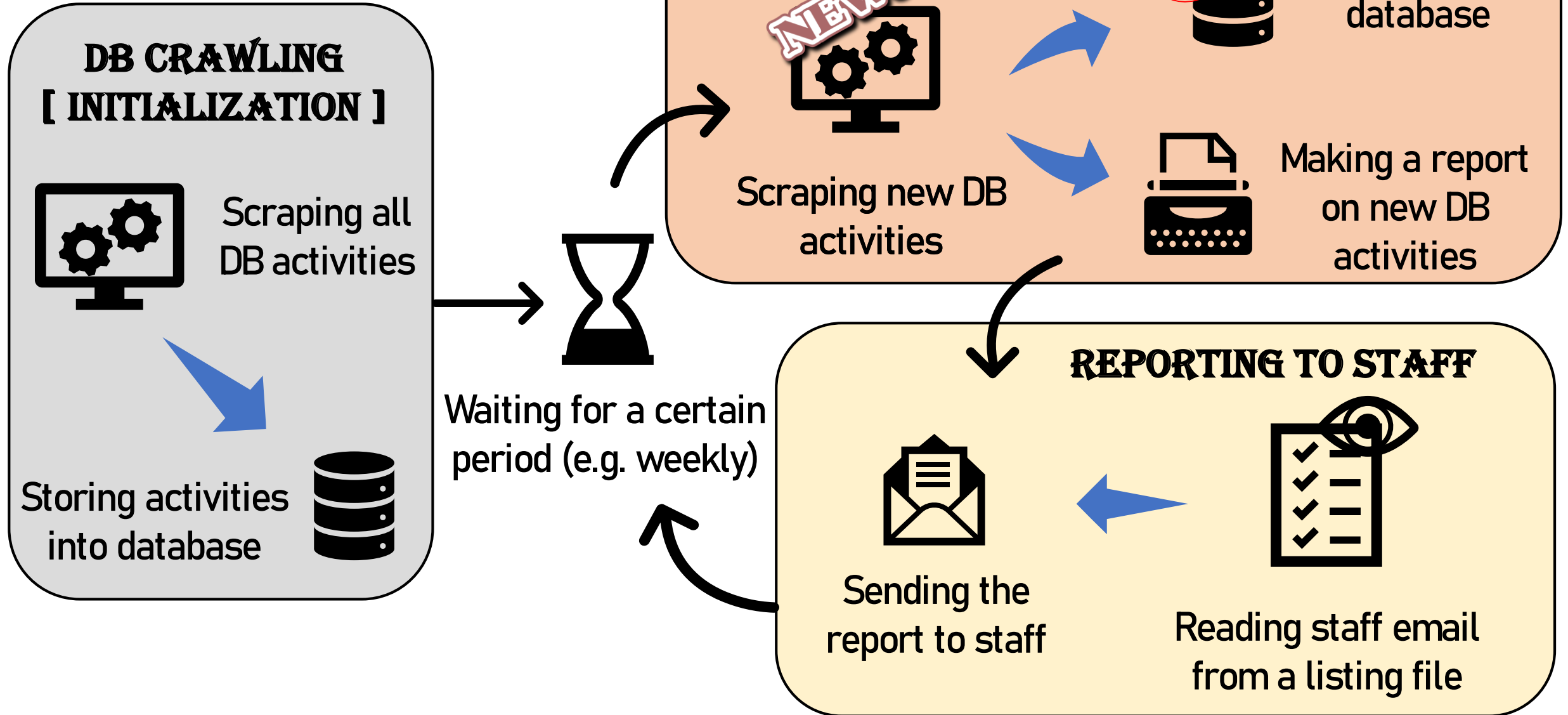
Developed by
Nopphon Keerativoranan
Teaching Assistant in OEDO

Last modified 14 October 2019

Objective

- To notify course staff and lecturers on the new activities in the MOOCs Tokyo Tech course via email address
- To store the discussion board textual data for further analysis

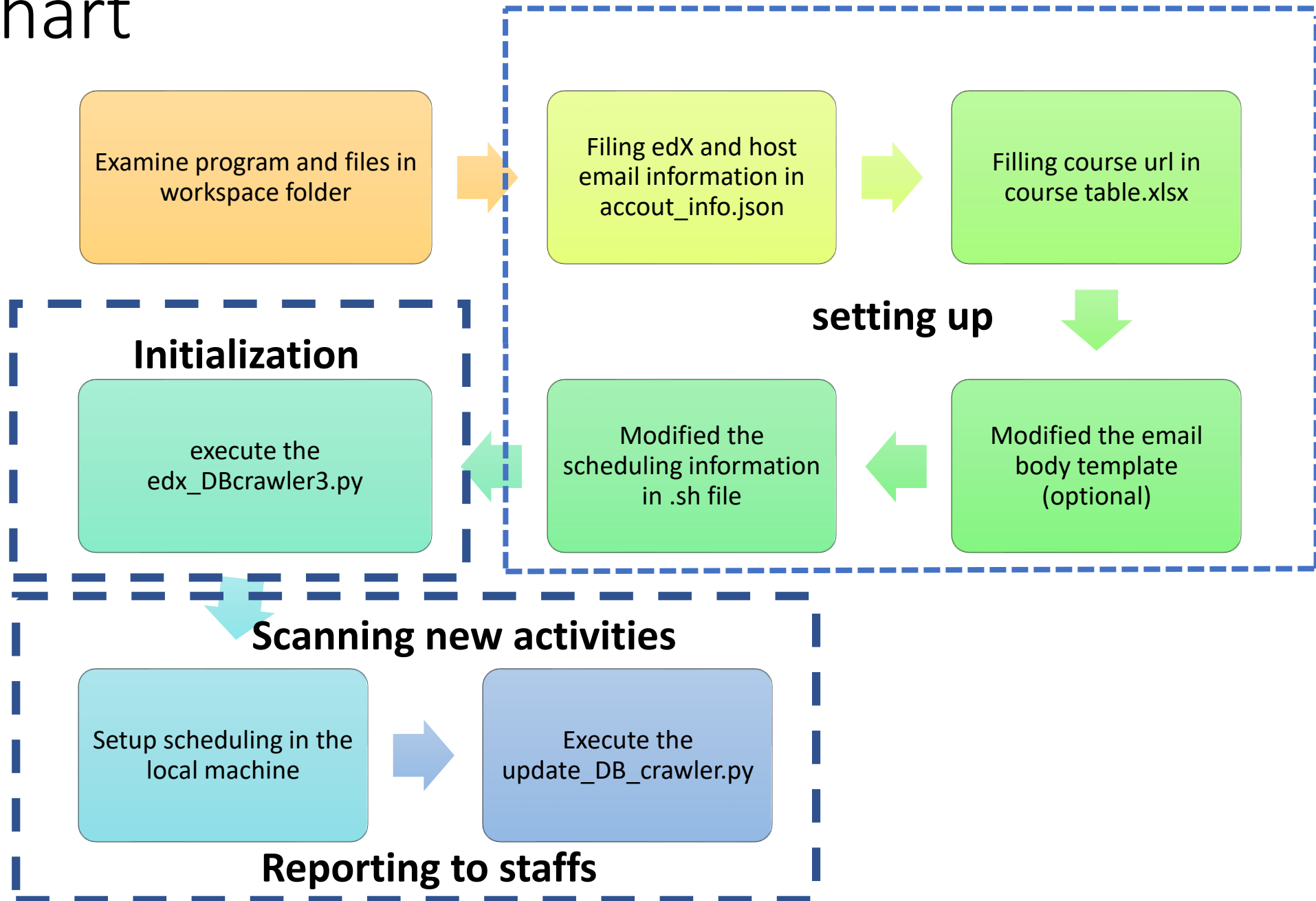
Overview



Prerequisite items

- **Dedicated edX account** for the detection the new activities in discussion board
- **Dedicated email address** as sender to send the report to desired recipients (Course staff, lecturer)
- **Dedicated Local machine with Ubuntu OS** to execute the program
- **‘new activities’** is defined as the posts, responses, or comments that the dedicated account has not been checked since the last scanning.

Flowchart



Workspace folder



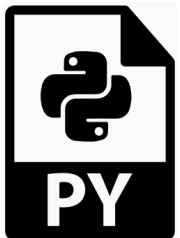
chromedriver

WebDriver for Chrome
<https://sites.google.com/a/chromium.org/chromedriver/>



Execute.sh

Bash script for task scheduling in Ubuntu



update_DB_crawler.py

Python script to crawl NEW DB activity and send email



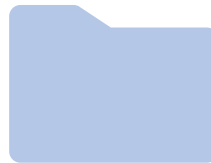
all_dis.json

Contains edX account and host email address



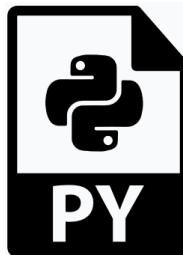
Course table.xlsx

Contains course URLs and recipients email address



HTMLs folder

Contains crawled data and a new activity report



edx_DBcrawler3.py

Python script to crawl all DB content

Additional information of Web Driver

- edX discussion board uses a dynamic web element
 - So, it needs a web browser to handle the dynamic Javascript

- Python-Selenium tool



(<https://www.seleniumhq.org/>)

- Provides Webdriver APIs of various browsers
- Can perform crawling at edX discussion board



Initialization

execute the edx_DBcrawler3.py

1) Execute the script with the following command

```
python edx_DBcrawler3.py
```

- Browser will be opened and goes to edX dashboard of the account indicated in *account_info.json*

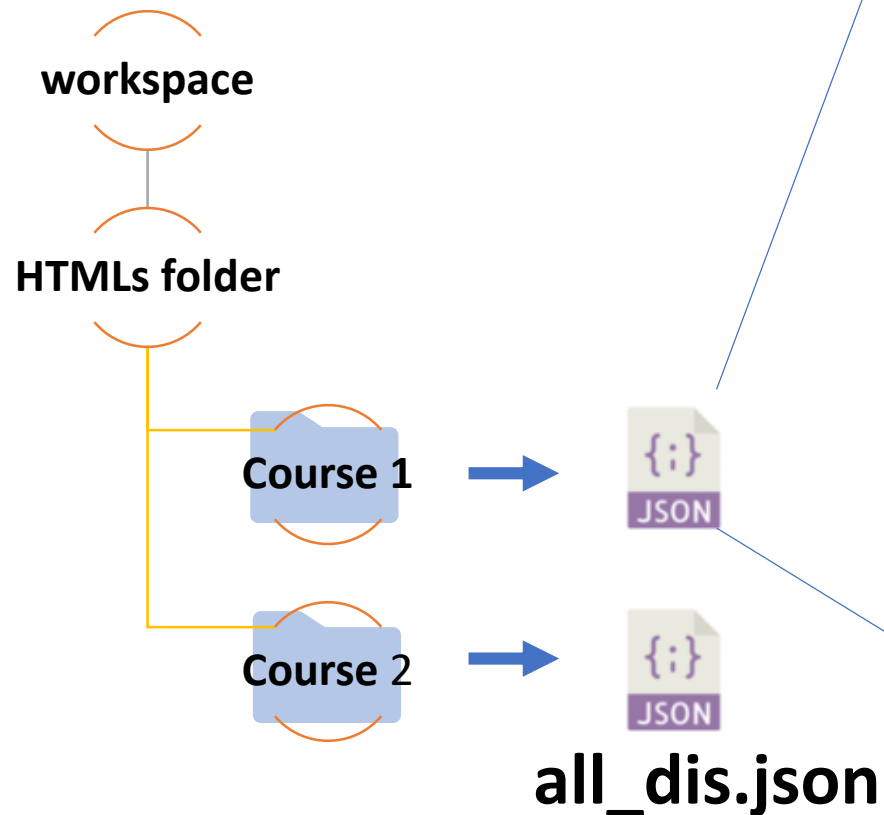
2) The crawler will ask which courses do you want to initially scrape the DB data

```
type 1 :select courses to crawl manually  
Type any keys : crawl all courses listed in excel file  
:
```

- Type your choice and the crawler will start scraping all DB data

Result

DB textual data is stored
in json-based database



all_dis.json

```
"0000": {
  "No_comment": 0,
  "No_response": 0,
  "post_category": [
    "0. Self-introduction",
    "Post here"
  ],
  "post_content": "Hi, I\u2019m a physio living in London who wants to find out more about health and complementary medicine with particular interest in physiology.",
  "post_timestamp": "2019-09-19T12:43:24Z",
  "post_user": "GTC75",
  "post_user_role": "n/a",
  "response": {},
  "response_role": {},
  "response_timestamp": {},
  "response_user": {},
  "title": "Hello",
  "type": "discussion"
},
"0001": {
  "No_comment": 0,
  "No_response": 0,
  "post_category": [
    "0. Self-introduction",
    "Post here"
  ],
  "post_content": "Hi! I\u2019m Arvin. I\u2019m a nurse, fitness instructor, personal trainer, massage and complementary therapist from London, United Kingdom. I\u2019m also doing a nutrition certification.",
  "post_timestamp": "2019-09-16T13:53:46Z",
  "post_user": "AR18",
  "post_user_role": "n/a",
  "response": {},
  "response_role": {}
}
```

Setting up

Account Information setup

account_info.json



```
{
  "edXaccount": "xxx@aaa.com",
  "edXpassword": "password1",
  "email_body_template": "message.txt",
  "email_body_error_template": "error_message.txt",
  "host_address": "yyy@bbb.com",
  "host_port": "587",
  "host_pwd": "password2",
  "host_smtp": "smtp.bbb.com"
}
```

Dedicated edX account and password for the query of new activities in discussion board

Body template files of email to be sent in case of tool running normally and abnormally

Information of the email sender

1. Email account
2. Password
3. Email protocol (usually smtp)
4. Email port (usually 587)

Adding course URL in the excel file

course table.xlsx

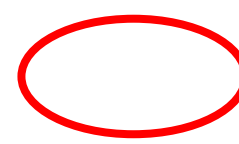


Course URL given from
the edX dashboard

course name	course url	directory	recipients (semicolon-separated)
Science and Engineering Ethics	https://courses.edx.org/courses/course-v1:TokyoTechX+Ethic1Jx+2T2017/course/	ETHICS 1	Prof. Micheal J. ,mcj@testemail.com,no; Course Staff, titechstaff@tetemail.com,yes
Introduction to Business Architecture	https://courses.edx.org/courses/course-v1:TokyoTechX+BA101x+3T2018/course/	BA 1	Prof. Steve B. ,stb@testemail.com,no; Course Staff, titechstaff@tetemail.com,yes

Make sure the url is
correctly added

Email body template [normal case]



Refer to the course names and recipients listed on course table.xlsx



Number of new activities found



\$DATE2 is the date of crawling

\$DATE1 is the date of previous crawling

(7 days ago by default. To change the period, please modify the date in update_DB_crawler.py)

Dear \$PERSON_NAME,

This is the discussion board activity of the TokyoTechX "\$COURSENAME" this week.

----- WEEKLY DISCUSSION BOARD NOTIFICATION -----

Period: \$DATE1 to \$DATE2

Course Name: \$COURSENAME

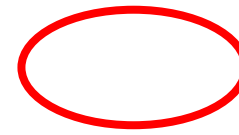
Number of new post/response: \$NEWPOST

Number of post with new comment: \$NEWCOMMENT

Please find attached file for further new post information.

Best regards,
OEDO team

Email body template [error case]



Refer to the course names and recipients listed on course table.xlsx



Number of new activities found



\$DATE2 is the date of crawling

\$DATE1 is the date of previous crawling

(7 days ago by default. To change the period, please modify the date in update_DB_crawler.py)

Dear \$PERSON_NAME,

There is an expected issue while retrieving the discussion board activity of the TokyoTechX "\$COURSENAME" this week.

----- WEEKLY DISCUSSION BOARD NOTIFICATION -----

Period: \$DATE1 to \$DATE2

Course Name: \$COURSENAME

----- ERROR LOG DETAIL -----

\$ERRORCONTENT

Best regards,
OEDO team



Error Log reported by the tool

Scanning new activities
&
reporting to staff

Adding directory and recipients information

course table.xlsx



Subdirectory name
inside 'HTMLs' folder

course name	course url	directory	recipients (semicolon-separated)
Science and Engineering Ethics	https://courses.edx.org/courses/course-v1:TokyoTechX+Ethic1Jx+2T2017/course/	ETHICS 1	Prof. Micheal J. ,mcj@testemail.com,no; Course Staff, titechstaff@tetemail.com,yes
Introduction to Business Architecture	https://courses.edx.org/courses/course-v1:TokyoTechX+BA101x+3T2018/course/	BA 1	Prof. Steve B. ,stb@testemail.com,no; Course Staff, titechstaff@tetemail.com,yes

You can freely change the directory name but make sure that it is matched with the name in this column

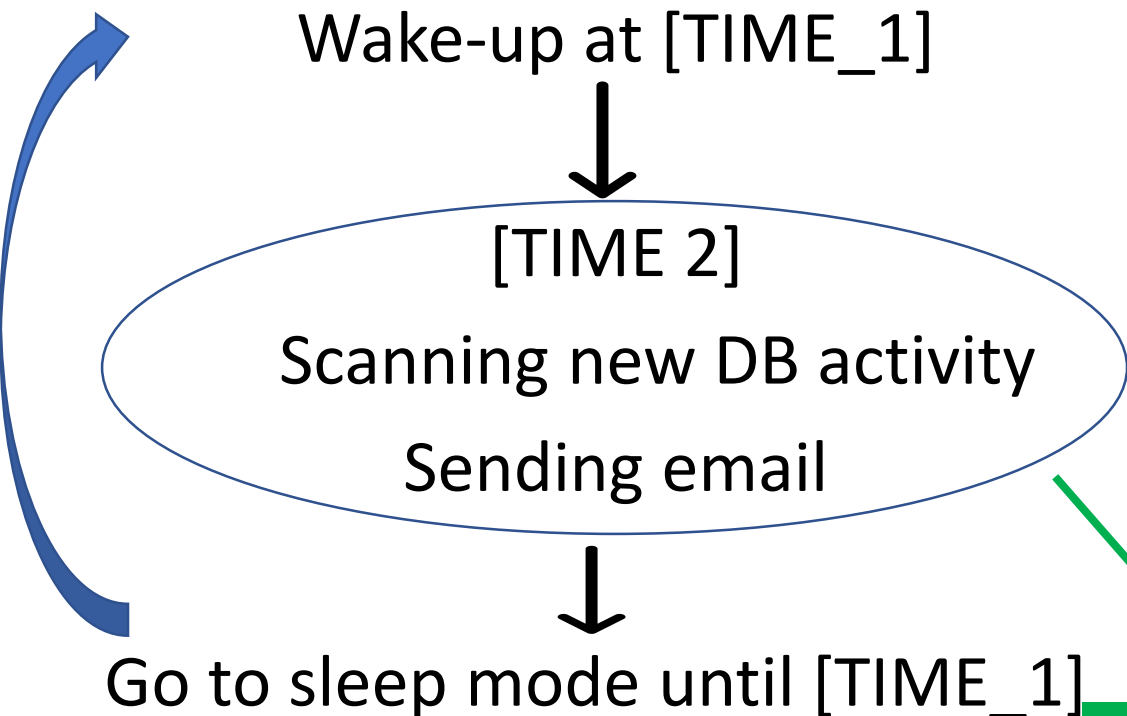
List of email recipients in the following format

[NAME_1], [EMAIL_1], [FLAG_1] ;
[NAME_2], [EMAIL_2], [FLAG_2] ;

*FLAG → if yes, send email when there are new activities
→ if no, send email even without new activity

Scheduling setup

The program will do 3 steps periodically regarding the right panel setup



Crontab based scheduling in Ubuntu OS

crontab -e

```
0 6 * * MON ./[WORKSPACE]/execute.sh
```

This line of code asks the OS to execute the execute.sh every [TIME_2] = Monday 6 AM

Your user's
pwd

Replace with your workspace directory.
Ex. [WORKSPACE] → home/auto-db/

execute.sh Script

```
#!/bin/bash

PATH=/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/[WORKSPACE DIR]

cd /[WORKSPACE DIR]/
echo 'ready to update' >> /tmp/crontest.log 2>&1
python update_DB_crawler.py >> /tmp/crontest.log 2>&1 &&

echo "PASSWORD" | sudo -S /usr/sbin/rtcwake -m disk -u -t $(date +%s -d 'next Monday 05:30') >> /tmp/crontest.log 2>&1
```

Logging file

execute the update_DB_crawler.py
file via execute.sh

Go to sleep and wake up at
[TIME_1] = Monday 5:30 AM

Result

New activities are
updated in database

Separated JSON file of new
activities are created

New activity report is created.

workspace
HTMLs folder

Course 1

Course 2

all_dis.json

newcomment,.json

newpost.json

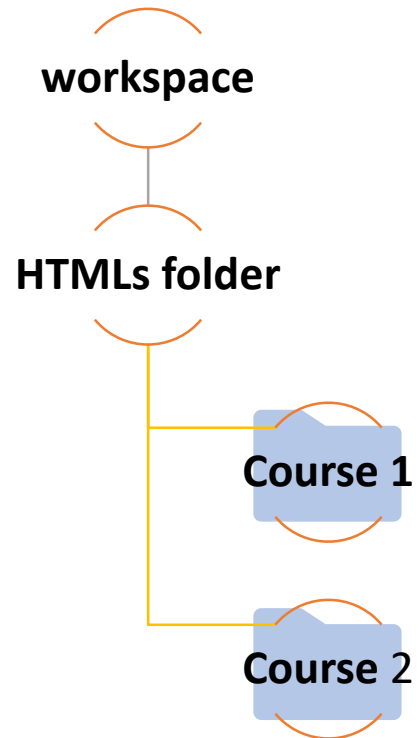
YYYYMMDD--DB_Notification—*COURSEDIR.xlsx*



Previous report is achieved in subdirectory



Notification report

[illegible]

Email example

Notification with new activities

Discussion board Notification: Autophagy ➤



oedotitech.db@gmail.com

to me ▼

Dear Nopphon,

This is the discussion board activity of the TokyoTechX "Autophagy" this week.

----- WEEKLY DISCUSSION BOARD NOTIFICATION -----

Period: 20191007 to 20191014

Course Name: Autophagy

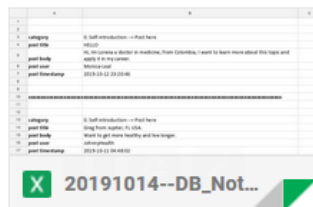
Number of new post/response: 20

Number of post with new comment: 0

Please find attached file for further new post information.

Best regards,

OEDO team



Report attached in excel format

Notification with NO new activity

Discussion board Notification: Basic Japanese Civil Law ➤



oedotitech.db@gmail.com

to me ▼

Dear Nopphon,

This is the discussion board activity of the TokyoTechX "Basic Japanese Civil Law" this week.

----- WEEKLY DISCUSSION BOARD NOTIFICATION -----

Period: 20191007 to 20191014

Course Name: Basic Japanese Civil Law

Number of new post/response: 0

Number of post with new comment: 0

Please find attached file for further new post information.

Best regards,

OEDO team

Person with FLAG = no will not receive this email

Adding a new course in the system



- Add course info in **course table.xlsx** according to slide 13 and 17
- Follow instruction **slide 9** but selecting **the 1st option** to select course manually
- Then **select only the new course** to be initially scrape all DB data
- DONE. In the next cycle, system will also check for the new course