**File System Crawler**

**Purpose**

Scan particular filesystem directory or directories. Tracks the files changes using a digest sum and last modification time. Stores the information in a database table. Provides a search utility to find all modified files for particular time range.

**Method**:

1. This project is divided into two modules. First module(file\_crawler\_v3.py) runs in the background and updates the database after every n scan-interval. Second module is a search functionality(main\_program.py) that searches the database based on time inputs by user as shown in the screenshot.
2. file\_crawler\_v3.py(file crawler version 3): This script runs in the background and take parameters from crontab and updates the database if any changes occur in the existing files(hash value changes) or Insert the data whenever new file is created.
3. Crontab: Crontab is used to schedule the script by passing required arguments needed for the script to run.
4. To create a crontab file run the following command on your command line:

export EDITOR=vim; crontab -e

              This command will open the vim editor and there you can specify time and parameters to  schedule the job as shown below

\*/20 \*/6 \* \* \* python /home/i871175/file\_crawler\_v3.py /home/i871175/prasunproject full

\*/20 \* \* \* \* python /home/i871175/file\_crawler\_v3.py /home/i871175/ top

The first command runs the script file\_crawler\_v3.py every 6 hours and 20 min. In this the parameters passed to the script are:-

Path: /home/i871175/prasunproject      type of scan: full

1. To run the main program file just type the file name followed by two inputs in the form of dates

                                     [i871175@usphlhanaags08 ~]$ python main\_program.py 2018-08-18 2018-08-22

DATABASE:

**Schema used:**

CREATE TABLE table\_fscrawler(fpath varchar(2048) primary key, fdigest nvarchar(32),fmtime datetime, rectime datetime);

Table Name: table\_fscrawler

Database Name: fscrawler