CSE207 Introduction to Networking
Project 2 - Report

Chen Qianshan

1508670

qianshan.chen15@student.xjtlu.edu.cn

Program: CST

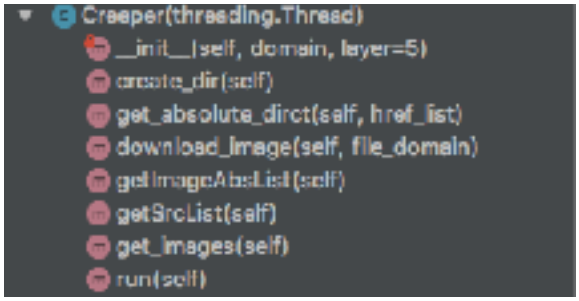
Lecturer: Charles Fleming

Due: Tuesday, 11:55 am, Dec.12 2017

Image Crawler Application.

A one-page report specifying which parts of the project you implemented, known bugs, and citing any sources you take more than 10 lines of code from.

1. This is an objective-orient program. The main class Creeper is inheritance from threading.Thread, a multi-thread build in package of python.
2. There are 8 methods in Creeper class. Shown in right. To initialize the class, **domain, and layer** are the two arguments domain is the aiming page waiting to be crawled, layer is the current layer, defaults as 5. To crawling each layer, the run() methods in the class will create and start a new Creeper class in one new thread. when the layer



3. Known Bugs:
   a. When crawling https://www.baidu.com the HTTP response body part cannot be decode with gzip or zlib.
   b. When the picture is big, it is possible to fail download it. Because of the socket connection timeout has been set to 2 seconds.
   c. User should enter the full https/http://xxx.xxx part for connection. When the web is a full ssl-site the program cannot automatically jump from http to https domain.

4. Citing source more than 10 lines
   a. Header part: These codes are get from the chrome browser to simulate a browser: