

Lab Test
Design Lab
April 5, 2023

We ask you to add following functionalities to your existing **select()** based chat-server assignment (Assignment 7) in the lab test:

1. Moderating hate-speech in the chat server: [30]

- a. Using a KNN-classification model trained on English hate/abusive-speech data ([link](#)), classify a sentence as hate/abusive or not.
- b. Then follow the below mentioned broad outline to identify hate-speech being sent as a message by some client:
 - i. Write a python script that takes a string input as a command line argument,, and using the classification model gives the label [**not_hate_abusive**, **hate_offensive**] and stores it in a temporary file (you can give it any name of your choice).
 - ii. Now invoke this python file as a subroutine (using `exec`, `execv`, `execp...`) from the server that passes the client's message as command line argument and retrieves the label from the temporary file you created.
 - iii. Then depending on the label take the action. If abusive: Notify to client —'Sent abusive message. !!Msg Blocked!!', else forward the message to the destination(s).

Functionalities to be modified: **/send, /broadcast, and /sendgroup**

2. Removing client that sends lots of abusive messages: [20]

- a. After identifying that client has sent an abusive message, for each client —maintain the count of instances in which the client has sent an abusive message. If the number of instances crosses a particular threshold (say **5**) remove the client from the server.
- b. Make sure that while removing the client you handle the following cases consistently:
 - i. Broadcasting all the other clients: '<Client-id> removed because of hate-speech'
 - ii. Dissolving the groups of which the client to be removed is the only admin.

Functionalities to be modified: **/send, /broadcast, and /sendgroup**