MosesSMT: User Guide

youtube.com/@proooooooo github.com/human71 cse23.xyz/materials

---

1. Install docker `$ sudo apt install docker.io`
   If you get a proxy error, run:
   `$ sudo mkdir -p /etc/systemd/system/docker.service.d`
   `$ sudo vi /etc/systemd/system/docker.service.d/http-proxy.conf`
   And paste the following lines:
   ```
   [Service]
   Environment="HTTP_PROXY=http://172.16.199.20:8080"
   Environment="HTTPS_PROXY=http://172.16.199.20:8080"
   ```
   After adding, run:
   `$ sudo systemctl daemon-reload`
   `$ sudo systemctl restart docker.service`

2. Pull the image using `$ docker pull techiaith/moses-smt`

3. Git clone `$ git clone https://github.com/porthtechnolegauiaith/moses-smt`

4. Get into the folder `$ cd moses-smt`

5. Run `$ make run` to start the docker image.

6. Till this point the cloned folder and the docker image should be connected, if not, run:
   `$ sudo make stop` and run `$ make run`

7. To check if the docker container is connected or not, make a folder *moses-models* into the repo you cloned, it should reflect in the docker *root* directory, if not, change line number 7 of *Makefile*, from *${PWD}* to the */home/your-username/moses-smt* and repeat step 5.

8. In the github repo, make a folder named *enhi* under the folder moses-models, and paste the dataset.

   ```
   |— moses-models
       |——— enhi
           |——— lm
           |——— corpus
               |————— enhi.en
               |————— enhi.hi
   ```

9. Go to the folder moses-smt in the docker shell and run the following for training:
   `$ python moses.py train -e enhi -s en -t hi`

10. Make a folder inside *moses-models* named *testing* and keep the text you want to test inside the src.txt file. Then, save the following code as `$test.sh` and run `$./test.sh`
    OR directly copy and paste the following code:

    ```
    while IFS= read -r line
    do
    "/home/moses/mosesdecoder/bin/moses" -f "/home/moses/moses-models/enhi/en-hi/engine/model/moses.ini"
    "$line" >> pred.txt
    done < "src.txt"
    ```

11. Run `$ ./multi-bleu.pl target_gold.txt < pred.txt` to evaluate your file using BLEU.