**Domain Specific Alignment Dictionary**

**About:** This module finds the meaning of the english words whose frequency is found by Termsuite

**Module/Task By:** Nupur Thakur (nupur.s.thakur@gmail.com)

**Documentation by:** Nupur Thakur(nupur.s.thakur@gmail.com)

**Last Updated**: 01/11/2019

**Prerequisite:** Termsuite Frequency based .tsv file for english words, Parallel-Corpus(Eng- Hin)

**Input: All the input files are inside Input Folder of the Module**

English Corpus : Merged\_eng.txt

Hindi corpus : Merged\_wx.txt

TSV Frequency File : AI\_TXT.tsv

**Output: All the output files get generated inside Output Folder of the Module**

1. Single word Dictionary - Single\_word.txt

2. Multi-word Dictionary - Multi\_word.txt

3. Unknown words whose meaning can’t be found - Unknown\_word\_list.txt

4. List of probable meaning for words - Probable\_dict.txt

**NOTE : To run the termsuite if you want to generate the list of frequent word for your corpus. Then, follow the instruction of the below readme.**

[**TERMSUITE\_README**](https://drive.google.com/open?id=12YLPHCXUPtxrq9ARvKpWaK7VT-3iQesorj-EmX8OhtE)

**Steps to Use the Module :**

1. Create Folder of the name Domain\_Specific\_Alignment\_Dictionary in Home Folder containing all the required files and Folders present in the module folder.

2. Copy the input files inside *Input* Folder. (i.e. .tsv file generated from Termsuite, English Corpus, Hindi Corpus)

**3. To run:** Before running the module make sure you are inside the Domain\_Specific\_Alignment\_Dictionary Folder.

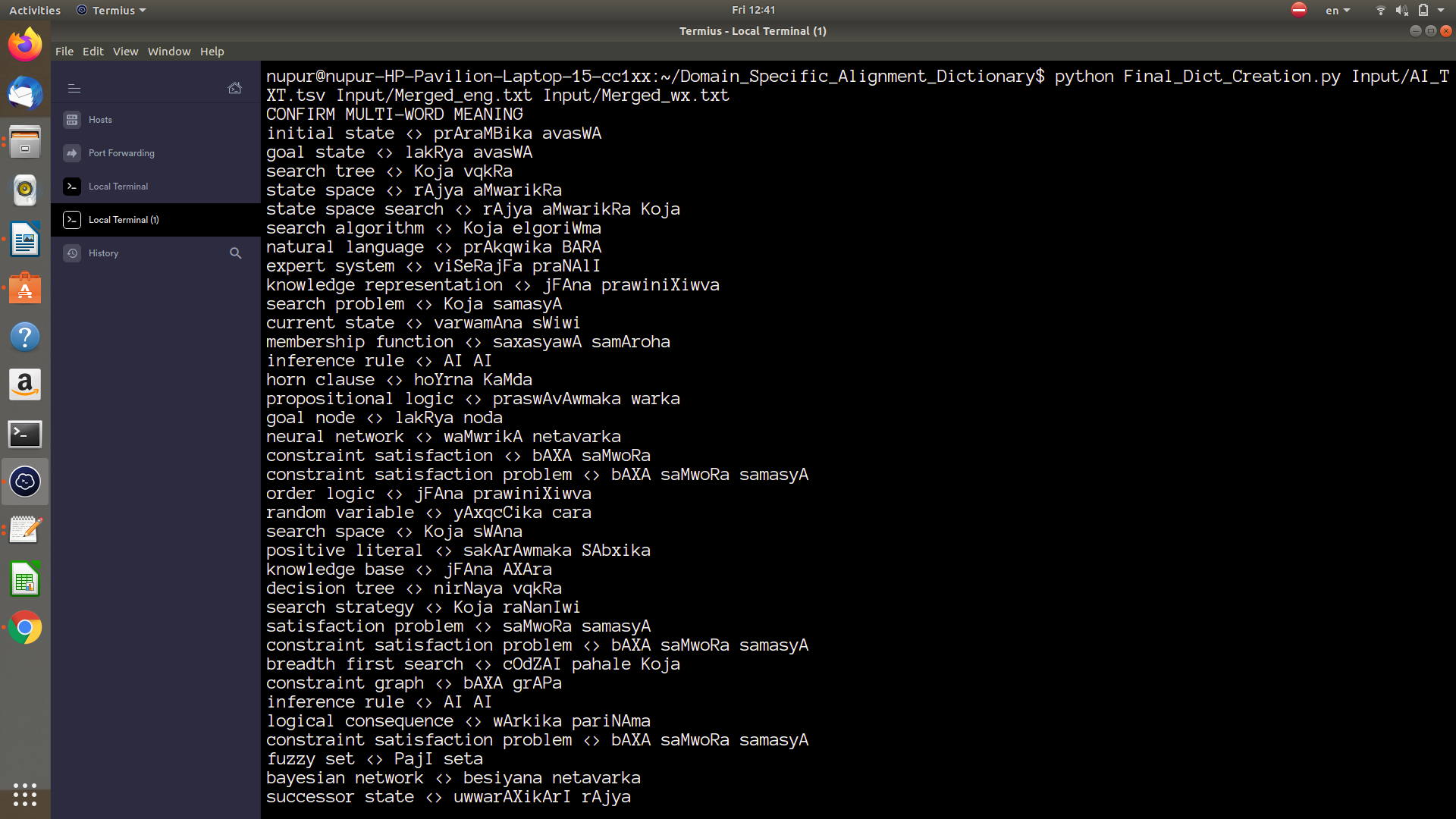
Then run the following command :

**python Final\_Dict\_Creation.py Input/<Termsuite\_generated\_file> Input/<english\_corpus> Input/<hindi\_corpus>**

**e.g.**

**python Final\_Dict\_Creation.py Input/AI\_TXT.tsv Input/Merged\_eng.txt Input/Merged\_wx.txt**

**(You can see the output generated inside Output Folder of the Module)**



Please submit in the form if you identify any errors/doubts etc in documentation - Link