

بسم الله الرحمن الرحيم

## پروژه ترجمه ماشینی (Machine translation) درس NLP

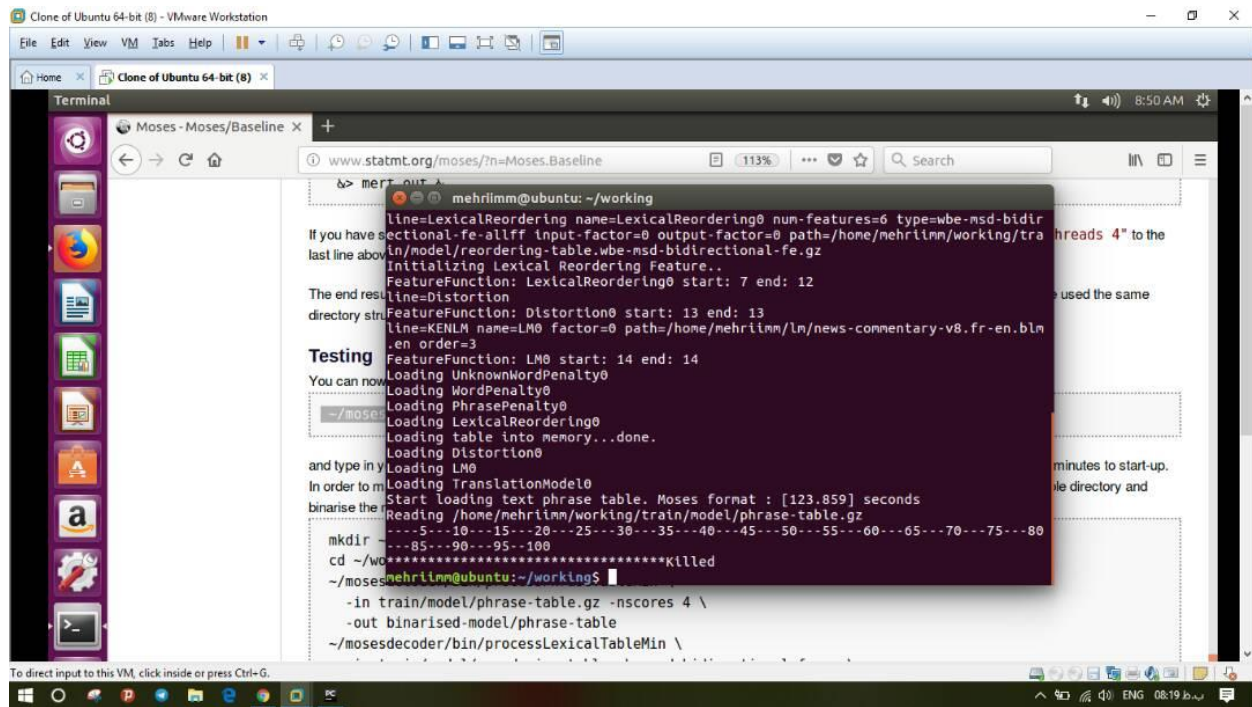
ابتدا متن کتاب "little women" را به عنوان داده، به زبان های انگلیسی و فارسی جمع آوری کردیم و هدفمان این بود که ماشین را با این داده ها آموزش دهیم و در آخر یک جمله به انگلیسی وارد کرده و بر اساس این آموزش ترجمه آن را به زبان فارسی بگیریم.

برای ترجمه ماشینی بعد از آماده کردن داده به دو زبان مختلف، باید مرحله alignment صورت گیرد. یعنی یک فایل حاوی متن داده به صورت دو به دو که ترجمه هر خط، در کنار آن قرار گیرد. سپس ماشین را با این فایل آموزش می دهیم طوری که در آخر یک جمله که از یک زبان خاص بدهیم، ماشین باید ترجمه آن را به آن یکی زبان بدهد.

برای آموزش از ابزار mooses استفاده نمودیم. بنابراین برای اینکار mooses را با توجه به لینک "<http://www.statmt.org/moses/?n=Development.GetStarted>" بر سیستم خود نصب نمودیم. سپس برای align کردن داده های دو زبان، از طریق لینک "<https://github.com/machinalis/yalign>" اقدام کردیم ولی متأسفانه با وجود دادن دیکشنری، برنامه در وسط کار خود killed می شد.

بعد از این تصمیم گرفتیم با توجه به مراحل لینک "<http://www.statmt.org/moses/?n=Moses.Baseline>"، به آموزش ماشین بپردازیم. در ابتدا GIZA++ را نصب نمودیم و مراحل بعد را نیز به راحتی گذراندیم و به مرحله training and translation رسیدیم که این مرحله تقریباً به مدت 2 ساعت به طول انجامید و بعد از گذشت این زمان داده های خود را tuning کردیم که این مرحله به مدت تقریباً 8 ساعت به طول انجامید. در آخر باید قسمت testing را انجام میدادیم که متأسفانه چندین بار سیستم در این مرحله هنگ کرد و با ریست کردن آن سعی در اجرای دوباره آن کردیم که در دفعات آخر بعد از گذشت حدوداً یک ساعت از پروسه، پیام killed آن به ما داده شد که تصویر آن در ادامه آمده است.

به نظر می رسد که این پروژه برای سیستم ما، به شدت سنگین بوده و طول کشیدن مراحل و انجام ناقص آن در مراحل مختلف به این دلیل می باشد.



سارا آئين

94521009