**NLP Project**

1. goal & topic

Creating a chatbot using Deep learning algorithm with movie subtitles

2. dataset from kaggle - Cornell Movie-Dialog Corpus

This corpus contains a metadata-rich collection of fictional conversations extracted from raw movie scripts:

- 220,579 conversational exchanges between 10,292 pairs of movie characters

- involves 9,035 characters from 617 movies

- in total 304,713 utterances

- movie metadata included:

- genres

- release year

- IMDB rating

- number of IMDB votes

- IMDB rating

- character metadata included:

- gender (for 3,774 characters)

- position on movie credits (3,321 characters)

3. task - try to use RNN and check the result

4. expected results - chatbot should be answerable

5. potential relevance for master thesis - Dealing with movie subtitles would be a challenging task. Data pre-processing is the most important task. For masters thesis it will give some ideas how to treat this kind of data.