Todo

1. Add more description of the way to get 92858 drug names and make it more clear.
2. The fair comparison of the result with the method CRF and HMM. Try to train the CRF and HMM with the semi-supervised labels.
3. From the table4, MRR value is low. However, the result indicated by table5 seems fine. Try to analyze or explain this difference.

In addition, better to choose the same drug in table4 and table5, and add more kind of drug instead of 3 or 4 kinds.

1. Enlarge the homemade dataset.

400 training data -> 1000 training data

200 test data -> 500 test data

1. (optional) Find the benchmark social media dataset on the web. Use the same dataset to get the result and compare the result with others.
2. Analyze the impact of step "homophone transform" and "extended ADR lexicon".
3. Analyze the importance of different features in step "training SVM classifier".
4. Modify some details.

Ex: reference, contraction, variable(italicized), figure.