

Bericht Masterpraktikum

Armin Bernstetter
University of Würzburg
September 24, 2019

ABSTRACT

This report is a summary of the

1 INTRODUCTION/MOTIVATION

- dataset erweitern
- language model

2 CRAWLER

- welche seiten kamen infrage (reddit, TA)
- wo lagen probleme beim crawlen?
- Was ist team andro
- wie ging der crawler vor

3 THE DATASET

- der bestehende super corpus
- Der neue Team Andro Corpus

4 ELMO

- Visualisierte elmo embeddings? TODO

5 GPT-2

- Welche implementierung und warum?
- der blog post mit den gedichten
- welche modelle kamen raus
- Fragen und Zweifel/Probleme. Undurchsichtige Trainingsskripte usw
- sample output und interactive prompt samples

6 WHAT COULD HAVE GONE BETTER? FUTURE WORK

- not depend so much on existing repositories/forks
- rather implement own training scripts for gpt2?