



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK



Bachelorarbeit

im Studiengang Computerlinguistik

an der Ludwig- Maximilians- Universität München

Fakultät für Sprach- und Literaturwissenschaften

Titel der Arbeit

vorgelegt von
Leonie Weißweiler

Betreuer:	Prof. Anna Korhonen
Prüfer:	Prof. Alexander Fraser
Bearbeitungszeitraum:	26. März - 28. Mai 2012

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, den 28. Mai 2018

.....
Leonie Weißweiler

Abstract

Dieses Dokument dient als Muster für eine Ausarbeitung einer Bachelorarbeit am CIS und wird in deutscher oder englischer Sprache erstellt (hier max. 250 Wörter)

Inhaltsverzeichnis

Abstract	I
1 Introduction	3
1.1 Ein Abschnitt	3
1.1.1 Ein Unterabschnitt	3
1.1.2 Noch ein Unterabschnitt	3
1.2 Beispiele	4
1.3 Mehr Beispiele	5
2 Introduction	7
3 Character-Level Recurrent Neural Networks	9
3.1 Recurrent Neural Networks	9
3.2 Long-Short Time Memory	9
3.3 Neural Architecture Search	9
4 Related Work	11
5 Datasets and Evaluation setup	13
5.1 The Wikipedia Corpus	13
5.2 Evaluation setup	13
5.3 Evaluation of RNN output quality	13
5.4 Evaluation of RNN productiveness	13
6 General Evaluation of Generated Text	15
7 Evaluation of RNN output quality	17
7.1 Overall Evaluation	17
7.2 Evaluation per Model	17
7.3 Evaluation per Language	17
7.4 Heap's Law approximation	17
8 Evaluation of RNN productiveness	19
8.1 Overall Evaluation	19
8.2 Evaluation per Model	19
8.3 Evaluation per Language	19
8.4 Heap's Law approximation	19
9 Conclusion and Future Work	21
9.1 Conclusion	21
9.2 The Future of Character-Level RNNs as language models	21
9.3 Character-Level RNNs for the sparsity problem	21
Literaturverzeichnis	23
Abbildungsverzeichnis	25
Tabellenverzeichnis	27

1 Introduction

[Schütze et al., 2008]

1.1 Ein Abschnitt

Mein Name ist Hase und ich weiß von nichts. Das ist ein Testtext. Mein Name ist Igel und ich weiß auch von nichts.

1.1.1 Ein Unterabschnitt

Blabla. Hier ein Unterabschnitt.

Ein Unterunterabschnitt

Blabla. Hier ein Unterunterabschnitt.

Noch ein Unterunterabschnitt

Wer 1.1.1 sagt, muss auch 1.1.1 sagen.

1.1.2 Noch ein Unterabschnitt

Das ist ein gewöhnlicher Absatz.

Ein Absatz mit Titel Paragraphen gibts auch.

Ein Unterabsatz mit Titel Und dann auch noch Unterparagraphen.

Ein nicht nummerierter Unterabschnitt

Dieser Unterabschnitt erscheint nicht im Inhaltsverzeichnis.

1.2 Beispiele

Blabla.

1.3 Mehr Beispiele

Blabla.

2 Introduction

[Schütze et al., 2008]

3 Character-Level Recurrent Neural Networks

3.1 Recurrent Neural Networks

3.2 Long-Short Time Memory

3.3 Neural Architecture Search

4 Related Work

5 Datasets and Evaluation setup

5.1 The Wikipedia Corpus

5.2 Evaluation setup

5.3 Evaluation of RNN output quality

5.4 Evaluation of RNN productiveness

6 General Evaluation of Generated Text

7 Evaluation of RNN output quality

7.1 Overall Evaluation

7.2 Evaluation per Model

7.3 Evaluation per Language

7.4 Heap's Law approximation

8 Evaluation of RNN productivity

8.1 Overall Evaluation

8.2 Evaluation per Model

8.3 Evaluation per Language

8.4 Heap's Law approximation

9 Conclusion and Future Work

9.1 Conclusion

9.2 The Future of Character-Level RNNs as language models

9.3 Character-Level RNNs for the sparsity problem

Literaturverzeichnis

[Schütze et al., 2008] Schütze, H., Manning, C. D., and Raghavan, P. (2008). *Introduction to information retrieval*, volume 39. Cambridge University Press.

Abbildungsverzeichnis

Tabellenverzeichnis

CD contents