

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
1.1	Motivation . . . . .	2
1.2	Problemstellung . . . . .	2
1.3	Aufbau der Arbeit . . . . .	2
<b>2</b>	<b>Daten und Vorverarbeitung</b>	<b>2</b>
2.1	Datenbasis . . . . .	2
2.2	Externe Programme und Hilfsmittel . . . . .	2
2.3	Vorverarbeitung . . . . .	2
<b>3</b>	<b>Word2Vec</b>	<b>2</b>
3.1	Parameter . . . . .	2
3.2	CBOW . . . . .	2
3.3	Skip-gram . . . . .	2
3.4	Negative sampling . . . . .	2
3.5	Hierarchical softmax . . . . .	2
3.6	Distanz zwischen Vektoren im Word2Vec Modell . . . . .	2
<b>4</b>	<b>Wikipedia-Korpus</b>	<b>2</b>
4.1	Gesamtkorpus . . . . .	2
4.2	Teilkorpus . . . . .	2
4.3	Testdaten . . . . .	2
4.4	Vergleich und Analyse . . . . .	2
<b>5</b>	<b>Spezieller Domänenkorpus</b>	<b>2</b>
5.1	Auswahl . . . . .	2
5.2	Crawling . . . . .	2
5.3	Testdaten . . . . .	2
<b>6</b>	<b>Vergleich Korpora</b>	<b>2</b>
<b>7</b>	<b>Zusammenfassung</b>	<b>2</b>
<b>8</b>	<b>Ausblick</b>	<b>2</b>
<b>9</b>	<b>Fazit</b>	<b>2</b>

# **1 Einleitung**

## **1.1 Motivation**

## **1.2 Problemstellung**

## **1.3 Aufbau der Arbeit**

# **2 Daten und Vorverarbeitung**

## **2.1 Datenbasis**

## **2.2 Externe Programme und Hilfsmittel**

## **2.3 Vorverarbeitung**

# **3 Word2Vec**

## **3.1 Parameter**

## **3.2 CBOW**

## **3.3 Skip-gram**

## **3.4 Negative sampling**

## **3.5 Hierarchical softmax**

## **3.6 Distanz zwischen Vektoren im Word2Vec Modell**

# **4 Wikipedia-Korpus**

## **4.1 Gesamtkorpus**

## **4.2 Teilkorpus**

## **4.3 Testdaten**

## **4.4 Vergleich und Analyse**

# **5 Spezieller Domänenkorpus**

## **5.1 Auswahl**

## **5.2 Crawling**

## **5.3 Testdaten**

# **6 Vergleich Korpora**

# **7 Zusammenfassung**

# **8 Ausblick**

# **9 Fazit**