

Inhaltsverzeichnis

1	Einleitung	2
1.1	Motivation	2
1.2	Problemstellung	2
1.3	Aufbau der Arbeit	2
2	Daten und Vorverarbeitung	2
2.1	Datenbasis	2
2.2	Externe Programme und Hilfsmittel	2
2.3	Vorverarbeitung	2
3	Word2Vec	2
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
4	Wikipedia-Korpus	2
4.1	Gesamtkorpus	2
4.2	Teilkorpus	2
4.3	Testdaten	2
4.4	Vergleich und Analyse	2
5	Spezieller Domänenkorpus	2
5.1	Auswahl	2
5.2	Crawling	2
5.3	Testdaten	2
6	Vergleich Korpora	2
7	Zusammenfassung	2
8	Ausblick	2
9	Fazit	2

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