

COS 314

NEURAL NETWORK

Project 3: Report

Author:

u11013878 - Jaco BEZUIDENHOUT

May 27, 2015

Contents

- 1 Dataset 1**
 - 1.1 Number of chunks used 1
 - 1.1.1 Afrikaans: 1
 - 1.2 How frequencies were calculated 1
 - 1.3 How special characters were handled 1
 - 1.4 Store format 1
 - 1.5 How the data was pre-processed 1

Chapter 1

Dataset

I have used a web scraper to scrape data from 2 websites:

- woes.co.za - Creative writing website for Afrikaans
- mibba.com - Creative writing in English

1.1 Number of chunks used

I used different size chunks. All were
↳ 300 characters.

1.1.1 Afrikaans:

1.2 How frequencies were calculated

1.3 How special characters were handled

1.4 Store format

1.5 How the data was pre-processed