Text Mining South Park

Dominik Nerger (i6146759)

June 2, 2017

Contents

1	Introduction	1
2	Data set	1
3	Pre-processing	2
4	Implementation and results	2
5	Conclusion	2
6	Future work	2

1 Introduction

This report is part of the second assignment for the course Computer Vision and focusses on face recognition using Principal Component Analysis and k-nearest neighbors classification.

2 Data set

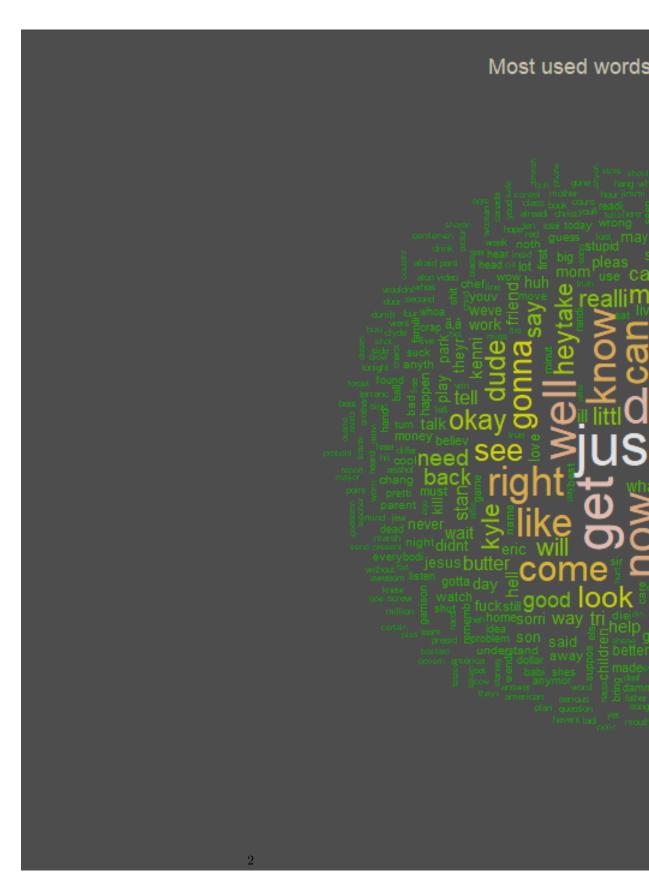


Figure 1: General wordcloud, containing terms by frequency

Words by tf-idf score over 1 manbearpig imaginationland jakovasau

Figure 2: General wordcloud, containing terms by TF-IDF score

Words by tf-idf score over 1 manbearpig imaginationland

jakovasau

Figure 3: Wordcloud containing ngrams (n=4) by frequency



Figure 4: Wordcloud containing ngrams (n=4) by frequency

- 3 Pre-processing
- 4 Implementation and results
- 5 Conclusion

In conclusion,

6 Future work