CS6370: Natural Language Processing

Spell Check Assignment Report

Indian Institute of Technology Madras

Chinni Chaitanya, EE13B072
ee13b072@smail.iitm.ac.in

Venkatesh Maligireddy, EE13B041
ee13b041@smail.iitm.ac.in

Swaroop Kotni, EE13B030
ee13b030@smail.iitm.ac.in

Pronnoy Noel, EE13B029
ee13b029@smail.iitm.ac.in

September 21, 2016

Contents

1 Introduction

This assignment attempts to implement *spell checker and correction* programs for text, which will correct the erroneous word[1] present in it. The assignment is categorized into the following three parts,

- **Word checker**, a program which checks and corrects standalone erroneous words[2].
- **Phrase checker**, a program which examines phrases and corrects the erroneous words. We assume *phrases* contain about 5 words.
- **Sentence checker**, a program which examines sentences and corrects the erroneous words. We assume *sentences* contain about 30 words.

In all the three cases, we assume that there is only **one** erroneous word.

2 Data generation

2.1 Corpa used for data generation

2.2 Data generated using these corpa

3 Word checker

..... Add content .....

3.1 Generation of candidates

3.2 Ranking the candidates using Bayesian probability

3.3 Ranking the candidates Phonetically

3.4 Combined ranking of candidates, using both Bayesian and Phonetic ranking

3.5 Observations

4 Phrase and checker

[1] In phrases and sentences, the erroneous word need not be misspelled but might be contextually incorrect.

[2] The word given might be correct word also.