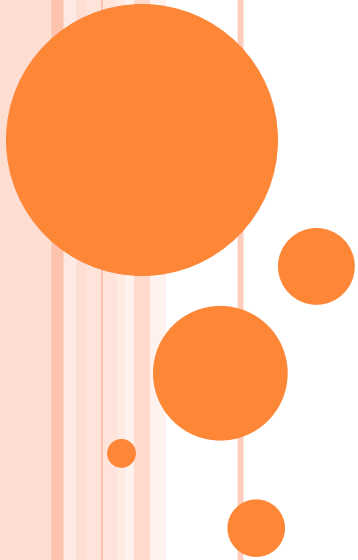


HINDI SPELL CHECKER

Amit Sharma, Pulkit Jain
Computer Science And Engineering,
IIT Kanpur-208016

Mentor

Dr. Amitabha Mukherjee
Computer Science And Engineering,
IIT Kanpur-208016



MOTIVATION

- Spell checking tools are important for editors, search engines etc.
- A lot of text is typed in Hindi
 - Books
 - Novels
 - Newspapers
 - Magazines
- Many spell checking tools exist for English, but not many for Hindi



INTRODUCTION

- Error Detection
 - Non Word Errors
 - Misspelled words are not part of the language
 - “ बन ” for “ वन ” (forest), “ दांत ” for “ दाँत ” (tooth)
 - Real Word Errors
 - Misspelled words are part of the language
 - “ दुकान उस और है ” for “ दुकान उस ओर है ”



INTRODUCTION..

- Correction
 - Find correction of the misspelled word
 - Find a correction c for word w such that

$P(c | w)$ is maximized

$$P(c | w) = P(w | c) P(c) / P(w)$$

- Produce a set of ranked corrections instead



INTRODUCTION..

- Ex : misspelled word = प्रमान
correct intended word = प्रमाण

```
>>> correct ('प्रमान')  
'प्रदान'  
>>> correct5 ('प्रमान')  
['प्रदान', 'प्रधान', 'प्रमाण', 'प्रमाणन', 'प्रमा']  
>>> |
```

The intended word is ranked 3rd and not 1st



PREVIOUS WORK

- Non Word Error
 - Dictionary Lookup
 - Word Frequency
 - Levenshtein - Damerau Edit Distance
 - **Most Widely Used**
 - N-Gram Analysis
 - Finite State Automatons
- Real Word Errors
 - Co-occurrence graphs
 - N-Gram Analysis



OUR GOAL

- Build a simple application
 - Allows user to enter text in Hindi
 - Rectifies misspelled errors in the entered text
 - Make use of the context to minimize real word errors in the text



REFERENCES

- [1] Tommi Pirinen and Krister Linden. *Finite-state spell-checking with weighted language and error models*. Proceedings of LREC 2010 Workshop on Creation and use of basic lexical resources for less-resourced languages [2010]
- [2] Francesco Bonchi, Ophir Frieder, Franco Maria Nardini, Fabrizio Silvestri and Hossein Vahabi. *Interactive and Context-Aware Tag Spell Check and Correction* [2012]
- [3] Suzan Verberne. *Context-sensitive spell checking based on word trigram probabilities* [2002]
- [4] Neha Gupta, Pratistha Mathur. *Spell Checking Techniques In NLP: A Survey* [2012]
- [5] Peter Norvig. How to write a spelling corrector. <http://norvig.com/spell-correct.html>

THANK YOU



QUESTIONS ?



LEVENSHTEIN DAMERAU EDIT DISTANCE

Number of edits required to convert one string to other.

Edits Include

- Splits
- Deletes
- Transposes
- Replacements
- Inserts





