

# Text Prediction using Probability Matrices and Prefix Trees

a CMSC190-2 presentation by  
John Alvin L. Sayson

# Significance of the Study

- Previous ventures into text generation allowed for applications in multiple fields.
- Writing mediums with embedded text prediction are now commonplace in mobile phones, with the goal of minimizing the general writing time in communication.
- The study aims to achieve a similar goal.
- Desktop systems will be given the power of predictive keyboards from mobile operating systems.

# Objectives of the Study

## GENERAL OBJECTIVE

The study aims to create an application that utilises similar functionality as a mobile prediction keyboard as an extension for the Google Chrome browser.

## SPECIFIC OBJECTIVES

1. Receive initial input from the user to instantiate the predictor application;
2. Use the probability matrix data structure in mapping word-level relation;
3. Use the prefix tree data structure in mapping letter-level relation;
4. Present word- and letter-level relation to the end user in the form of a predictive text writer interface;
5. Allow user text selection to create texts and allow said texts to update previously mentioned data structures; and
6. Assess user experience and program outputs using surveys.

# Methodology

# Implementation

- Javascript
- Google Chrome API library for browser integration

# Text Input Representations

- Sequence
- Bag-of-words
- Probability Matrix
- Prefix Tree

# Application Specifications

- Input - accept preliminary input from an HTML page
- Processing - clean input and convert to 4 representations, and store said representations to local storage
- Output - form results depending on level of prediction, display result in a table of choices, determine word-level probabilities using Bayes' theorem formulas

# Results

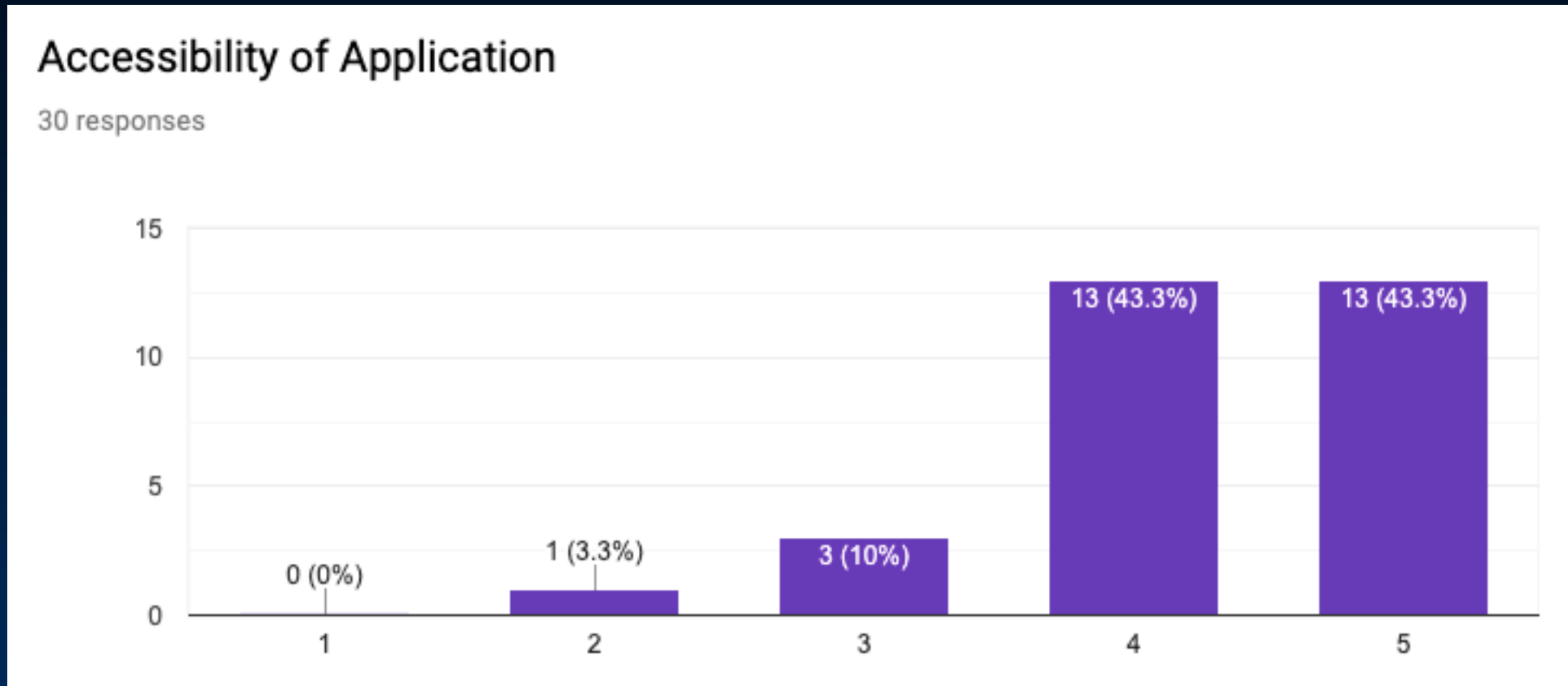


Fig. 1. Accessibility of application results



## Word-Level Prediction Accuracy

30 responses

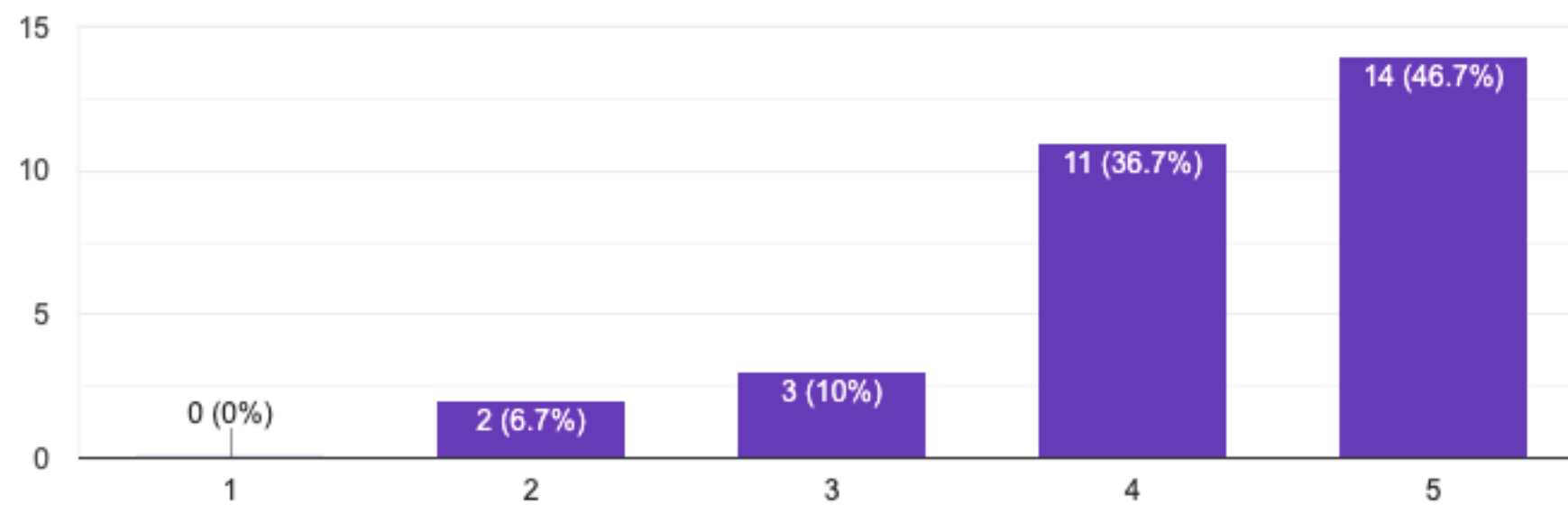


Fig. 2. Word-level prediction accuracy results

## Letter-Level Prediction Accuracy

30 responses

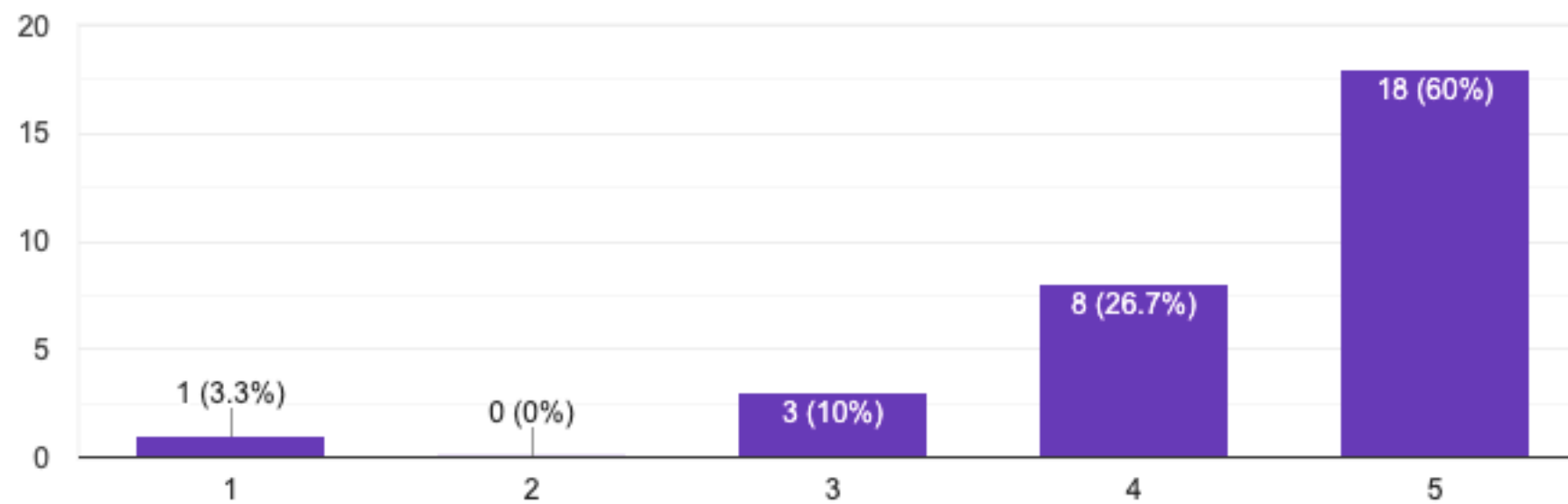


Fig. 3. Letter-level prediction results.

## Application Robustness

30 responses

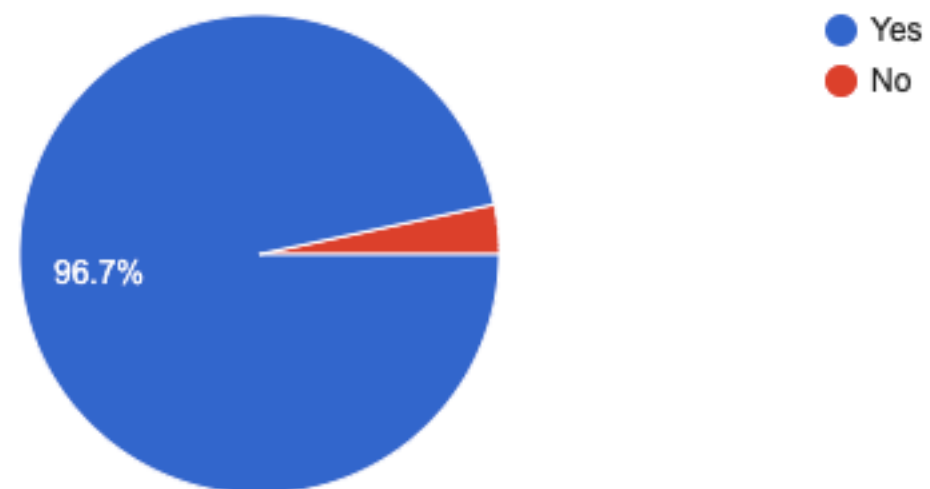


Fig. 4. Application robustness results.

## Overall User Experience

30 responses

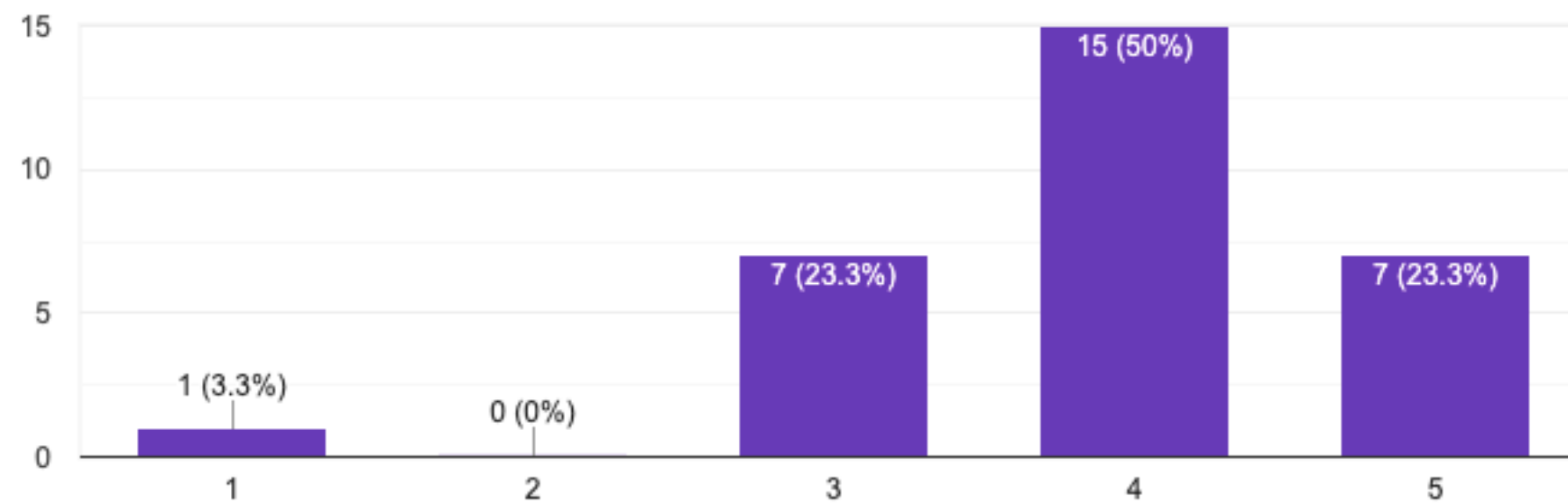


Fig. 5. Overall user experience results.

## Overall Application Rating

30 responses

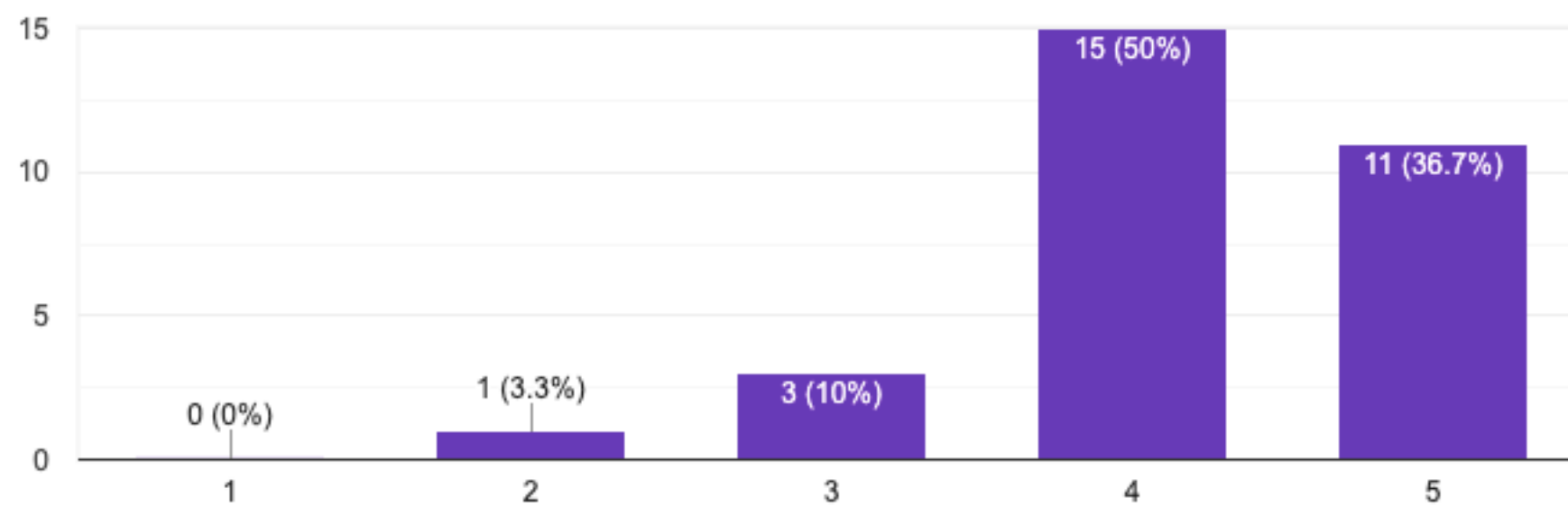


Fig. 6. Overall application rating results.