Eye-typing Software: Predictive Typing System
Sponsors: WSU School of EECS, Team Gleason
Mentor: Jesse Waite
By: Andre Gomez, Casey Jamrog, Joseph Niles, Tom Carrothers

Background
In terms of eye-typing, every click counts. For our project we created software that predicts the current and next word that the user is trying to type. Every correct prediction decreases the amount of time the user will have to click, thus increasing typing speed and providing a better experience for the user.

Project Goal
The goal for our project is to provide a competitive prediction engine to be integrated with the Team Gleason eye-typing user interface. The software will be open-source, which will be made freely available. The target user is people with ALS, where eye-typing systems are their only form of communicating with others.

Software Features
- Auto-Completion of current word
- Prediction of next word to be typed
- Dynamically updating the word database
- Parsing of personal emails and text files

Prediction Model (N-gram)
- N-gram is a list of items which can be either words or letters that have been taken from an arrangement of text or speech corpus.
- Our prediction software utilizes the COCA corpus, which contains two, three, four, and five n-gram models of millions of words.
- Given these models, we can make a prediction of the next word the user is trying to type.

Evaluation Metrics
1. Keystroke reduction - the amount of keystrokes it takes to type using the prediction as opposed to typing the full word
2. Total prediction accuracy.
   - Our goal for the reduction of keystrokes was 30%.
   - Our goal for prediction accuracy was 25%.

Data Structures
- Trie Tree (Next Word Prediction):
  - A tree data structure in which all leaves of the tree are arrays of items.
  - The Trie tree holds each gram of the COCA corpus, in which each word is a node in the tree and the predictions for the word are the leaves.
  - In terms of the structure, the predictions will be the words with the highest frequency values according to the input before it.
- Ternary Tree (Auto-Completion):
  - A tree data structure similar to the Trie, only each leaf of the tree holds a single character and a frequency.
  - For every character, the path will generate several choices of words containing the highest frequency given the entered characters.
- SQLite Database:
  - A serverless, self-contained SQL database engine.
  - Stores data in the filesystem.
  - Each database contains four tables, each with three columns of data.
    - Previous Words
    - Next Word
    - Frequency

Interpolated Trigram Model:
\[ P(W_n | W_{n-2}, W_{n-1}) = \lambda_1 P(W_n | W_{n-2}) + \lambda_2 P(W_n | W_{n-1}) + \lambda_3 P(W_n) \]
Where: \( \sum \lambda_i = 1 \)

Analysis
- Graph 1 shows the percentage of keystroke reduction over time using only the general database.
- Graph 2 shows the prediction accuracy of the custom database, COCA database, Auto-Completion, Combined results and the target accuracy.

Future Work
- Further development of merging algorithm between general database and custom database.
- Add dynamic weights to N-gram prediction model.
- Develop current context based prediction vs just user history.

Broader Impact
- Integrating this prediction software with the Team Gleason eye-typing interface provides people with ALS a competitive product with no cost, giving users the ability to communicate at a comfortable pace and higher quality of life.

Glossary
- ALS - Amyotrophic Lateral Sclerosis; a neuromuscular disease which inhibits muscle function
- COCA - Corpus Of Contemporary American english; a corpus containing several n-gram frequency tables used in our algorithms

Special Thanks To:
Jesse Waite, Gail Gleason, Sakire Arslan Ay, Dave Bakken, Aaron Crandall, and Jon Campbell

Team Force