CS 411 Final Project Report

1. Project Changes

In our initial proposal we proposed using an intelligent speech recognition and pronunciation analysis feature to allow users to practice speaking phrases and get real-time feedback on their pronunciation. However, we had to make changes to this to align with the goals of the course and focus on implementing and testing database systems rather than an Artificial Intelligence system. Our updated proposal now focuses on implementing a scoring system for sentences based on their length, where users are rewarded for answering challenging questions correctly. We also correctly addressed the concerns with scraping the data off the wikivoyage website and processed the data to make it usable.

2. Project Application & Utility

With this project, our goal was to help users to equip themselves to survive in a country that speaks an unknown language. Our main focus was helping tourists rather than dedicated learners that truly wanted to master a language. We believe this goal was achieved as not only did we focus on teaching essential phrases for basic communication, we also integrated features that tourists would be interested in by providing the popular monuments and dishes of the local cuisine in the 10 most popular cities for the language they picked. With these features, we aim to provide a complete experience for our users to give them an all-inclusive platform to complement their trip. An element that we think we could focus on is improving the overall functionality which was infeasible as this project was done only as part of a semester-long project. This project has immense potential to become the go-to platform for tourists and provide additional features such as finding hotels, book tourist attractions & restaurant reservations.

3. Data Source & Schema

We used Wikivoyage to source our phrases in English, French, Spanish, German, and Portuguese and generated synthetic data for the monuments and popular dishes which were then verified to ensure authenticity and the validity of the dataset. We also made changes to our initial schema to include the monuments and dishes tables. We also created a table for the user-friend relations.

4. UML Diagram & Table Implementations

In our initial UML Diagram, we had a separate entity for each language and a users and friends entity. We had to change the language entities as it was an inefficient way to handle them with separate entities. Instead we changed it to one combined phrase entity. Additionally, the friends entity was changed to a self-loop on users (implemented as a table), and we added two more entities for monuments and dishes to add more features to our application. Our final design was far more correct and suitable than the original design with more functionality as well.

5. Advanced Database Programs

In our application, we believe our advanced database queries enhance user experience by helping them learn about the countries related to their languages and connect with other people. Two queries help users discover dishes and monuments and the corresponding cities and countries for the language they are learning, so they can get excited about learning the language. We used an advanced query to display the leaderboard which ranks the friends of each user based on the points. Finally, another query helps users meet mutual friends who study the same language.

6. Technical Challenges

The first issue we encountered was getting GCP setup as none of us had used the service before. While initially surveying the features it seemed quite intimidating but with the help of the workshops, we were able to set up our accounts and host our database. Another issue we had was learning the basics of javascript to understand how

to code the frontend and integrate it with our database and backend. To overcome this, we used Youtube tutorials along with ChatGPT to create a frontend that is well-integrated with our databases. Rutva also ran into a specific issue when trying to upload the monuments and dishes datasets which was caused due to the encoding type. By default the data was of UTF-8 encoding which led to the text being weirdly represented due to the special characters present in languages like French. Our solution to this was converting the encoding type to iso 8859-1 which has support for Western languages and hence correctly represented the data on our frontend interface. Syed encountered an issue while designing a trigger to update points dynamically when adding or removing a point. Initially, the trigger failed to function as expected, and he realized it was because the logic didn't include updating the points directly within the trigger. Once he identified and corrected this oversight, the trigger performed as intended.

7. Future Work

As mentioned above, a lot of the improvements to the project would be adding more functionality as it is currently limited due to the time constraints. Besides helping with travel such as hotels and restaurants mentioned above, other features would include better training exercises, which could include different data modalities, more languages to study, and other types of information such as important facts on culture and norms.

8. Division of Labor

Labor was divided fairly. Teamwork was managed well through meetings and online communications. A lot of work was done together in meetings, which was equal. In terms of individual work on code, everyone contributed their own advanced queries. Brendan worked on code for preprocessing, cleaning, and preparing data from the web source. Pranav worked on creating the sql files and setting up the database on GCP. Rutva worked on connecting the front end and back end. Syed worked on implementing

the transactions, store procedures, and triggers. All of us got together and worked on completing the front end and adding all the features.