FindYourRestaurant: Find Flavor, Wherever You Are.

Kevin Heinrich, 11902941

Matthias Hemmer, 11804194

Christina Knes, 11902902

Fiona Trapp, 11903030

Institute of Interactive Systems and Data Science

Course: 706705 Advanced Information Retrieval

November 12, 2024

Design Document

Title

FindYourRestaurant

Group Number

Group 14

Members + Roles and responsibilities

- Fiona Trapp:
  - Frontend
  - Other on-demand tasks (Backend, dataset)
- Christina Knes:
  - Dataset processing and documentation
  - Other on-demand tasks (Backend)
- Kevin Heinrich:
  - Backend
  - Other on-demand tasks (dataset)
- Matthias Hemmer:
  - REST API
  - Backend

Idea

The research question is formulated as follows: How can advanced information retrieval techniques improve personalized restaurant recommendations by incorporating user preferences, contextual factors, and sentiment analysis?

To begin with, the user should provide some input via a UI in the form of natural language. In addition, it will also be possible to select certain filters (e.g. pet-friendly, family-friendly) from a dropdown menu. The user query as well as the selected filter options are then further processed by making a dedicated API call.

The API is responsible for communicating with the frontend and backend. For the former, it has to serve the data retrieved from the latter.

The backend should represent the heart-piece of the whole project. It contains the sentiment search model that is used to evaluate and rank restaurants according to their relevance to the user. Decisions made are based on the restaurant data that is stored in a SQL database.

Main Task

The main task is to create an IR system using the yelp restaurant data and word2vec. This system should then, based on the provided user query, output restaurant recommendations ranked according to their relevance.

Dataset

The dataset in use can be found at https://www.yelp.com/dataset/documentation/main. It consists of yelp business, reviews and user data, structured as JSON files.

Methods/Models + Data Processing

After the required data is stored in the SQL database, text from restaurant customer comments and ratings are pre-processed, before being fed into the sentiment search model. Pre-processing of the data includes tasks such as the removal of special characters and the conversion to lowercase. After that, the texts are tokenized, word embeddings are calculated and vectorized.

For each restaurant, the model shall output a relevance score indicating the meaningfulness with respect to the provided user query.

To train the sentiment search model, the k-fold cross validation strategy is used. This strategy divides the data into k random samples, one of which serves for validating the results. In order to find the best value for k, the cross validation has to be performed with different values (i.e. setting $k = 2, 3, 4, \ldots$). Depending on the training and validation accuracy of the results, the parameter k is selected.

Model parameters of the sentiment search model are fine-tuned via backpropagation over multiple epochs. Regarding the number of hidden layers and neurons per layer, a decision will be made at a later development stage.

Evaluation

As already mentioned above, k-fold cross validation is applied. To evaluate the performance in this case, the metric results for each fold are stored and analyzed to determine the overall performance and variability.

Visual Depiction

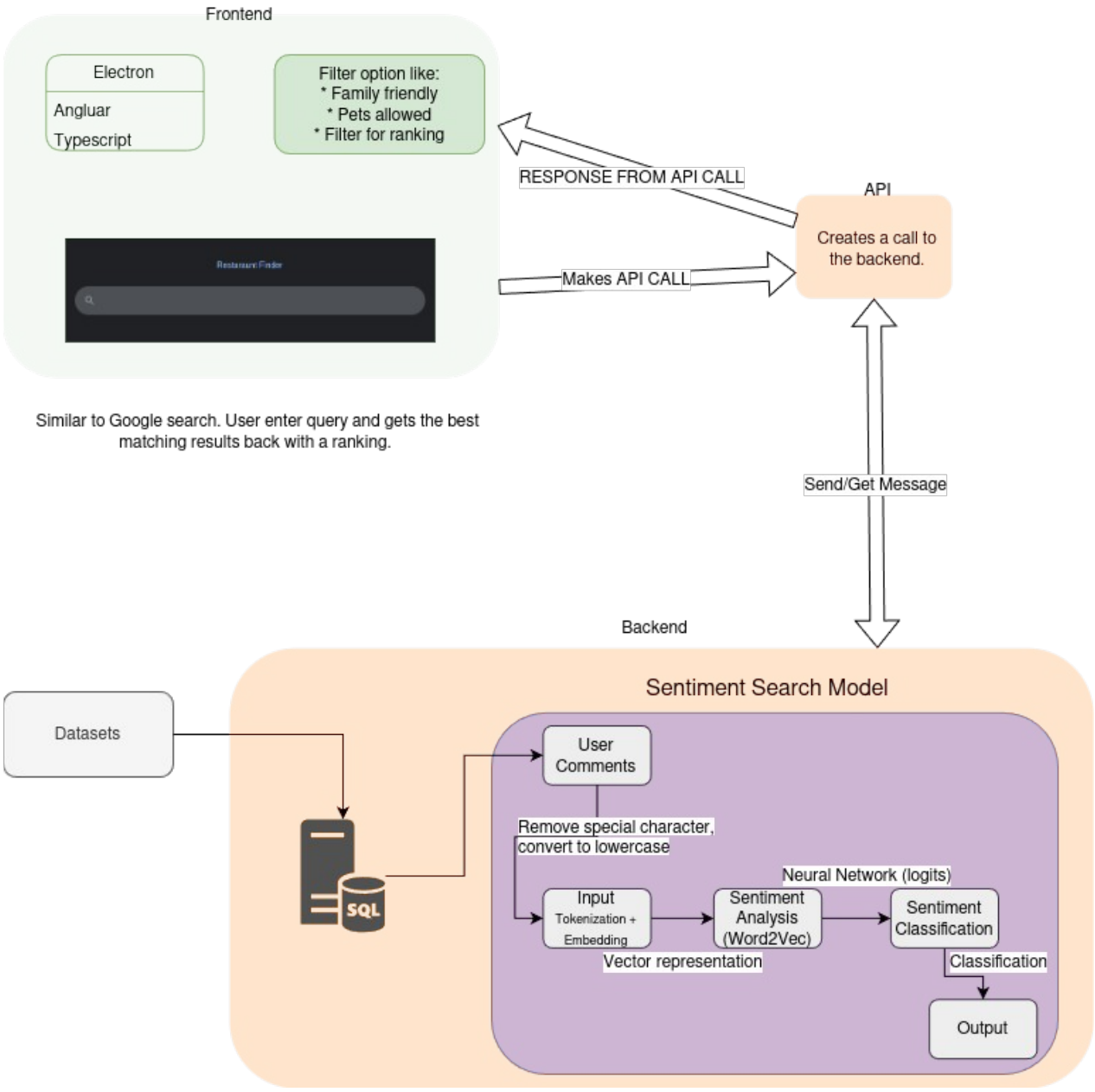

Figure 1: System Architecture