Development of a Virtual Cooking Assistant

Hugo Mojica Angarita, Camila Rincón Rojas
Universidad Distrital FJC
Bogotá, Colombia
hmojicaa@udistrital.edu.co, jcrinconr@udistrital.edu.co

***Abstract*—This paper presents the development of a chatbot designed to classify and separate information contained in a recipe dataset, focusing on the division of ingredients and instructions. The main goal is to demonstrate how an automated system can process and organize large volumes of culinary data, making them more accessible and usable. The relevance of the topic lies in the growing demand for digital tools that optimize information management in the culinary field. The applied methodology includes the use of natural language processing (NLP) techniques to analyze and categorize the data. The results demonstrate the effectiveness of the chatbot in identifying and separating the essential components of each recipe, thereby improving user experience when interacting with the database.**

Keywords: Chatbot, natural language processing, recipe classification, ingredients, instructions.

I. Introduction

The digitalization of culinary information has enabled the creation of vast databases with detailed recipes. However, efficient management of this information remains a challenge. This work presents the development of a chatbot that uses natural language processing techniques to divide and organize the information contained in a recipe dataset, facilitating its consultation and utilization.

A. Research Topic

Development of digital tools for managing culinary data.

B. Study Object

Recipe dataset with ingredients and instructions.

C. Research Motivations

Optimization of culinary information management and accessibility.

D. Topic Relevance

Increase in the use of digital technologies in cooking and the need for efficient tools.

E. Research Problem Background

Existence of large volumes of culinary data without a clear structure to facilitate their use.

F. Purpose

Demonstrate the effectiveness of a chatbot in classifying culinary data.

G. General Objective

Develop a system to separate and organize ingredients and instructions from recipes.

H. Specific Objectives

1) Implement NLP techniques in the chatbot.
2) Evaluate the system's accuracy.
3) Improve user experience in recipe querying.

II. Methodology

The methodology of this work was divided into several key sections:

A. Subjects

The dataset used includes a diverse collection of recipes with details about ingredients and preparation instructions.

B. Techniques and Instruments

- **Data Loading:** PDF files are loaded using `PyPDFDirectoryLoader`, which extracts the textual content from PDF files in a specified directory.
- **Segmentation:** The extracted text is divided into segments using `RecursiveCharacterTextSplitter`, with a segment size of 10,000 characters and an overlap of 20 characters.
- **Embeddings:** Semantic embeddings are generated for the text segments using the `sentence-transformers/all-MiniLM-L6-v2` model from HuggingFace.
- **Vectorization:** The embeddings of the segments are stored in a FAISS index to facilitate fast and efficient retrieval.

C. Procedure

- **Problem Formulation:** Identification of the need to separate and organize culinary data.
- **Definition of Objectives:** Clarification of the chatbot's objectives.
- **Information Search:** Relevant datasets and PDFs were searched about the topic.
- **Chatbot Development:** Implementation of NLP techniques to process the dataset, using `LlamaCpp` as the language model.
- **Evaluation:** Tests to measure the chatbot's accuracy in classifying ingredients and instructions and generating responses to user queries.

III. Results

The results show that the chatbot is capable of effectively identifying and separating ingredients from instructions. In preliminary tests, several recipes were evaluated to measure the system's accuracy and precision. The results indicated that, on average, the chatbot could correctly classify the ingredients and instructions of the analyzed recipes.

A. Low-Calorie Recipe Analysis

In evaluating the caloric content of the provided recipes, the following recipe was identified as the least caloric option. The table below shows the details of this recipe.

TABLE I

You: Tell me the least calorie recipe
Assistant: The least caloric recipe here is the Shredded wheat biscuit. With a total of 102 calories for one serving, it's the best option if you are looking to minimize your caloric intake while still enjoying a tasty and satisfying meal or snack.

This recipe was selected as the option with the lowest caloric content available, making it ideal for those looking to minimize their caloric intake while enjoying a tasty and satisfying meal or snack.

B. Other Notable Results

In addition to the least caloric recipe, further analyses were conducted to evaluate the accuracy and effectiveness of the chatbot in classifying ingredients and instructions. Graphs and tables were used to visually represent the accuracy distribution across different types of recipes. Table II presents questions and results showing the chatbot's accuracy in classifying recipes into various categories.

TABLE II

You: Recipe for people with fatty liver
Assistant: I'm sorry, but the provided recipe contains chicken livers, which are high in cholesterol and saturated fats, making it not an ideal choice for individuals with fatty liver or those trying to avoid high cholesterol foods. Instead, consider opting for other sources of protein that are lower in cholesterol and saturated fats, such as beans, lentils, or fish. Additionally, including a variety of fruits and vegetables in your diet can help support overall liver health.
You: What is the healthiest recipe
Assistant: The title of the recipe doesn't provide enough context to determine its healthiness. In general, a healthy recipe might include whole grains, plenty of vegetables and fruits, lean protein sources, and healthy fats. However, it ultimately depends on individual dietary needs and preferences.

IV. Discussion

The main findings indicate that the use of NLP techniques is effective for classifying culinary data. The discussion addresses the implications of these results in improving information management in culinary databases. The chatbot's ability to enhance the accessibility and usability of large volumes of culinary data is highlighted, offering precise and organized responses to user queries. Furthermore, possible improvements are proposed, such as incorporating more training data and fine-tuning the language model, as well as future research to further optimize the system and expand its capabilities.

References

[1] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media, 2009.

[2] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson, 2019.

[3] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[4] Niharika, *Nutrition details for most common foods*, 2019. [Online]. Available: https://www.kaggle.com/datasets/niharika41298/nutrition-details-for-most-common-foods

[5] PES12017000148, *Food ingredients and recipe dataset with images*, 2019. [Online]. Available: https://www.kaggle.com/datasets/pes12017000148/food-ingredients-and-recipe-dataset-with-images