nourish

A Knowledge-Driven Recommender System for Food Entrepreneurs

UC San Diego
JACOBS SCHOOL OF ENGINEERING
MAS DSE - Capstone Project

Abstract

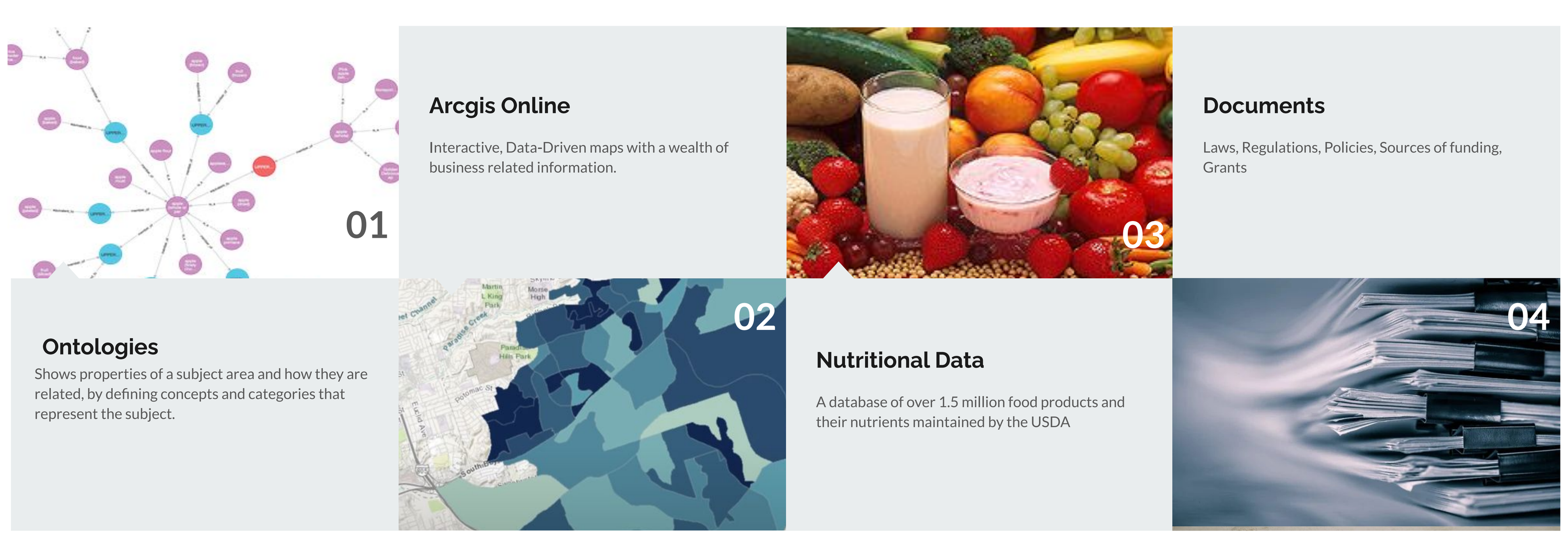
The food industry faces a significant knowledge management challenge. Professionals working in this industry must navigate the complexities of identifying potential markets, managing supply chains, adhering to laws, policies, and regulations, and securing funding, all while ensuring the survival of their businesses. However, the required information is dispersed across numerous disconnected data systems, posing a significant hurdle in accessing relevant and crucial insights.

Furthermore, professionals often lack the necessary means to effectively search across these disparate data systems, further exacerbating the problem. This fragmented information landscape leads to knowledge gaps that have far-reaching negative consequences for food-related businesses. These repercussions include an abundance of unhealthy food options, limited business opportunities, marginalized communities, the existence of food deserts, and the prevalence of food swamps.

Addressing this problem necessitates the development of a cohesive solution that consolidates the fragmented data systems and provides professionals with an efficient means of searching for the information they need. By closing these knowledge gaps, we can foster healthier food options, expand business prospects, empower marginalized communities, and combat the issues of food deserts and food swamps.

Acquisition

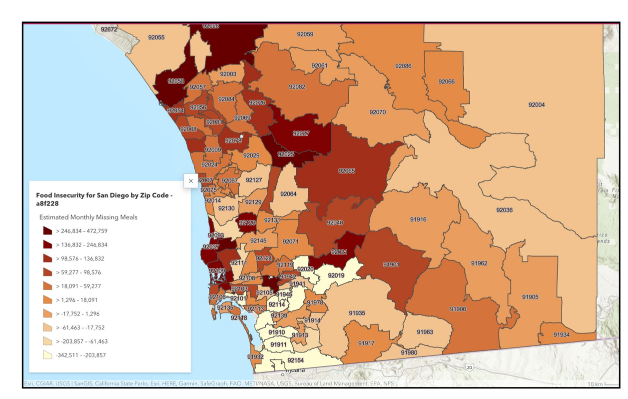
To establish a comprehensive source of domain knowledge, a knowledge graph was constructed between the ontologies and the USDA data. The Food Ontology and USDA food products were linked using an open-source GitHub repository called LexMapr. The various ontologies obtained from Ontobee were merged into a single OWL file using Protege, an open-source ontology editor. Subsequently, a custom script utilizing OwlReady2 parsed the class information and relationships from the OWL file. The output of this script was a valid Neo4j node and relationship file.



Scaling & Reporting

The Funding Document Tool is built by leveraging the Llama index module to create indexes of the available SBA and USDA funding documents. Documents are stored on a Google Drive folder and an API is used to access the directory of files. LlamaIndex then used to creates embeddings on the document store by chunking the documents. Llama Index then queries these embeddings by taking the user input and running a string similarity comparison to help find relevant documents. Once the document chunks have been selected, the information is passed to GPT to reason over the response. Llama index provides an invaluable service by dynamically reducing the amount of textually context needed for GPT to provide a relevant answer.

Preparation & Analysis



To further enhance the Nourish application's domain knowledge, the ArcGis data was transformed and structured. There were well over five thousand demographic and business variables which made feature selection inevitable. Through manual selection, the most relevant features were extracted from the data, grouped, then stored in feature layers.

Since one of the primary goals of the Nourish application is to recommend where a user should start or expand a business, a method for suggesting locations with the highest opportunity was needed. In order to do so, an opportunity score was created for each census block group as a high overview summary. The opportunity score was based on eight values.

- Demographics
- Population Density
- Foot Traffic
- Competition
- Opportunity
- Affordability
- Crime

Each of these values were created via feature extraction using ESRI variables. Each variable is normalized based on other block-group values and centered around 0. Then a linear optimization function is made as below:

*a*Demo+b*PopDen+ c*Foot Traf-d*Comp + e*Opp+ f*Aff-g*Crime + h*GrowthPot*

Category	%kcal from fat	%kcal from carbohydrates	%kcal from simple sugars	%sodium by weight
FSOD: Fat and Sodium	>25%	-	-	≥0.30%
FS: Fat and Simple Sugars	>20%	-	>20%	-
CSOD: Carbohydrates and Sodium	-	>40%	-	≥0.20%

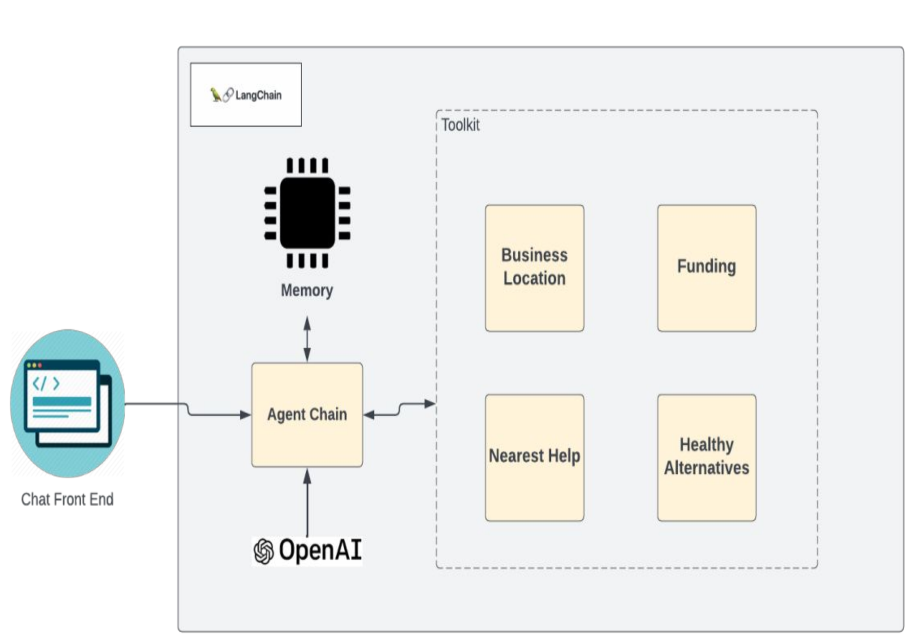
A table was created indicating the true/false value for the inclusion of each cluster for every food item.

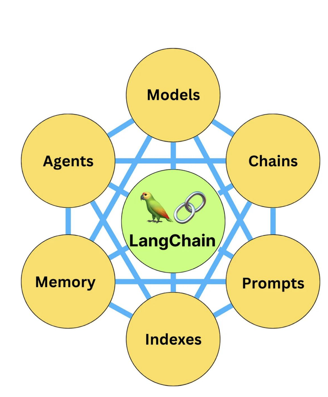
Since Nourish also aims to promote healthy food-related businesses, a method for suggesting the healthiest alternatives for various types of foods was necessary. Similar to the opportunity score, each USDA food product was assigned a nutrition score. However, determining what foods are considered healthy is often subjective. To address this, a method that focuses on identifying hyperpalatable foods was chosen. Hyperpalatable foods typically consist of processed foods or sweets with enticing combinations of fat, sugar, carbohydrates, and sodium.

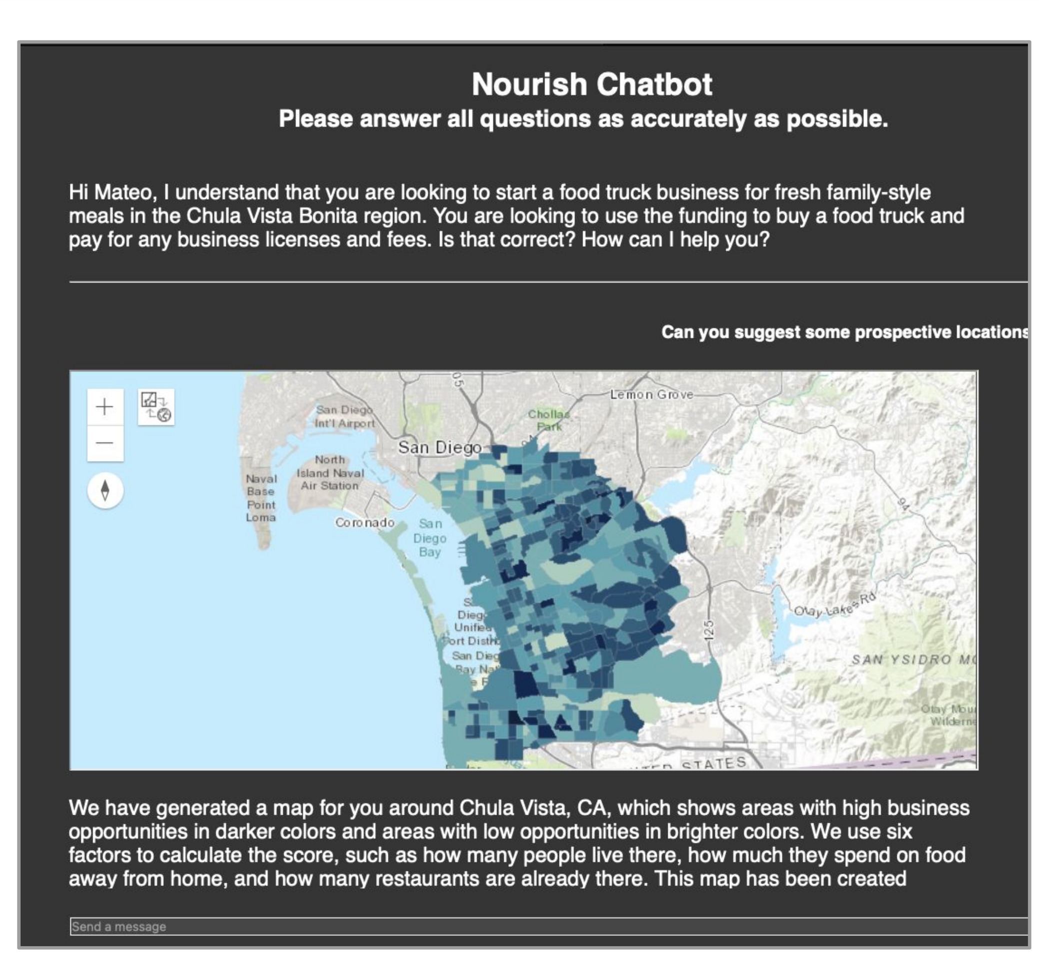
To establish a comprehensive source of domain knowledge, a knowledge graph was constructed between the ontologies and the USDA data. The Food Ontology and USDA food products were linked using an open-source GitHub repository called LexMapr. The various ontologies obtained from Ontobee were merged into a single OWL file using Protege, an open-source ontology editor. Subsequently, a custom script utilizing OwlReady2 parsed the class information and relationships from the OWL file. The output of this script was a valid Neo4j node and relationship file.

Final Solution

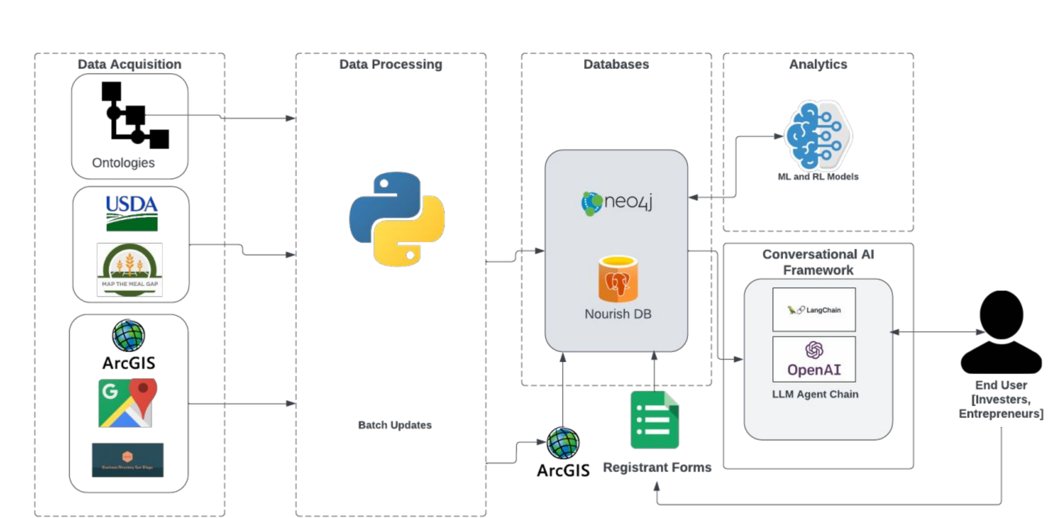
Along with its conversational capabilities, Langhain also provides a valuable memory feature. This feature allows our AI agent to retain and update the entire conversation as it unfolds during a chat session.







Dash, Flask, and a web based front delivers a simple interface. It displays the questions and answers given by the LLM and allows the user to type responses. Similar to interacting with the user, Nourish has to be trained to interact correctly with the knowledge base. To achieve this, a toolkit was developed. Each "tool" in the toolkit can be considered as a recommendation tool. The included tools are funding, locations, healthy alternatives, and nearest help.

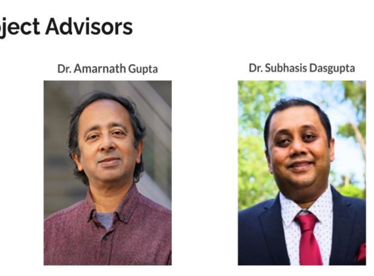


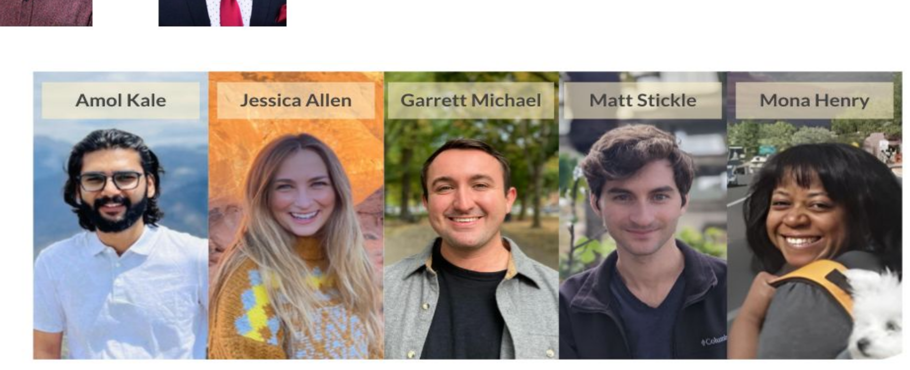
Key Insights

- AI Technology is evolving rapidly which means the Nourish application will require frequent evaluations for enhancements.
- We are able to demonstrate the viability of a recommendation system capable of handling a wide range of topics.
- Given the proper restrictions, an LLM can act as a liaison between different sources of knowledge.

Meet the Team

Project Advisors





Acknowledgments

The Nourish Project was funded by grants from the NSF.

Special thanks to: PI Laura Schmidt, Dr. Ilya Zaslavsky, and Dr. Tera Fazzino

National Science Foundation