

Christine Trinh

CS 4395.001

March 11, 2023

## Building a Corpus

### **How did I build the knowledge base?**

I built a knowledge base by recursively web scraping links and their content. I had to do a lot of filtering of extraneous data via preprocessing and frequency analysis. After reading in about 400 links, I prepared raw text that I wanted to be processed. Then, I went through each and made note of words I would see the most. After going through the most common ones, I handpicked about 10 terms to build a knowledge base about. I did this by simply saving sentences containing that term into a dictionary.

### **What are my top 10 terms?**

1. cuisine
2. history
3. Levantine
4. Indian
5. Arbëreshë
6. North
7. American
8. Greek
9. Italian
10. Albanian

## What does my knowledge base look like?

indian

It literally translates to "Indian dates", which is the Arabic name for tamarind.

It also incorporates elements of larger French, Indian and Chinese cuisines, brought by French colonialization and Indian and Chinese immigrants respectively.

[25] The cuisine of Martinique is a Creole cuisine with a mix of French, indigenous, African, and Indian cooking styles using local ingredients such as breadfruit, cassava, and christophine.

However, if referring to Eurasian cuisine, usually the amalgamation of Portuguese, Dutch, British, Chinese, Malay, and also Indian and Peranakan influences is meant.

arbëreshë

[citation needed] Kanojët is a typical Sicilian pastry and very common among the Arbëreshë people, which brought that dish back into their homelands, Albania where it is popular[citation needed].

The kanojët from Piana degli Albanesi, an Arbëreshë village, are often referred to be the best cannolo.

[citation needed] Other dishes include mëlci pule, eggplant appetizers, panaret which is famous among Arbëreshës, stuffed peppers composed of green peppers stuffed with rice, meat, other vegetables and herbs, turshi lakre, fried sardele me Limon, papare, which are bread leftovers cooked with water, egg, butter, and Gjizë (salted curd cheese) and bread and cheese referred as Bukë me djathë.

The Arbëreshë cuisine (Albanian: Kuzhina Arbëreshë; Italian: Cucina Arbëreshë) is the cuisine of the Arbëreshë people in Italy.

## How could I use this knowledge base to build a chatbot?

My chatbot could match as many terms as possible to the query and output the sentence. A sample dialog is as detailed.

Chatbot: Hello, I am Babish. I know all there is to know about Mediterranean food. What would you like to know?

How does Mediterranean food taste?

It has a range of tastes ranging from bright (think lemon, parsley, vinegar) to warm (cumin, pepper.) There is bound to be something you would like.

Is Mediterranean food healthy?

There is research to suggest that following a Mediterranean diet can aid in the prevention of cardiovascular diseases, and with increasing lifespan. When combined with a caloric deficit, weight loss can even be a result.

Are there any Mediterranean restaurants near me?

I found a few: Luna Grill, Istanbul Grill, Al Rayan Kabab & Gyros, Chop House