University of CINCINNATI

Dr. Chia Yung Han
Faculty Advisor

Ben Hollar
Computer Science

Erin Graska
Computer Science

Stephanie Tam
Computer Science

The SpiceRack

Project Abstract

Finding, collecting, and cooking recipes can be a daunting task. There are numerous hurdles for people to effectively manage their recipes. The Spice Rack aims to make the following effortless for users:

1. Parsing a variety of sources of recipes into a common, easy-to-read format.
2. Storing user recipes, including automatically imported recipes from the our parser.
3. Presenting an effective user-interface for navigating stored recipes.

Design Overview

Recipe URL → ML Extraction → Web App

The application begins by taking a URL to a webpage that has a recipe. We then perform two steps of machine learning (ML)-driven parsing of that URL:

1. Extract text content that belongs to the recipe (remove ads, blog text, etc.)
2. Classify extracted text into the structure of a recipe (title, ingredients, instructions)

Once the recipe is parsed, we can display the result to the user along with the rest of their collection.
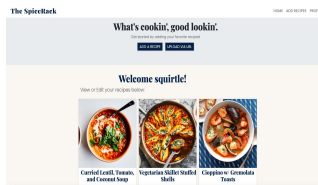
Home Page Layout with interactive cards that allow viewing/editing recipes

Recipe card display with buttons to edit & delete recipes
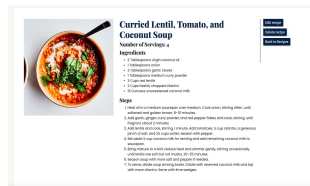
Takeaways

Ben: Creating an effective machine learning model is all about your dataset; I wish I'd known sooner how much time it'd take to compile even a modest set of recipes to train on. I'm happy with our results, though.

Erin: Making sure to have all functionality in mind when designing the database is very important, so that way when getting data in the future it's easy to do so. And to also have even stretch goals in mind so the database doesn't have to change when new functionality is added.

Stephanie: Figuring out how a website is laid out is not that easy. I'm pretty picky when it comes to aesthetics, so for me it was hard to decide how the content should be laid out. Bootstrap was my savior. In the future I will definitely do some wireframing before I try to put together the UI.

Obstacles

Machine Learning Results

While our content extraction machine learning model achieves an F1 score (a combined metric of precision and recall) of 0.83, it still frequently misses content of the recipe or incorrectly identifies content that is not part of the recipe.
This could be overcome by curating a larger training dataset.

Interface Design

Creating an effective, modern user interface is a difficult challenge even for practiced corporations with large development teams and budgets. We leveraged a web-development library, Django, to take care of common concerns and enable us to prototype faster.

Future Possibilities

Optical Character Recognition

Eventually, we would like to have an option for users to incorporate their traditional handwritten recipes by uploading an image of their recipe, allowing a machine learning algorithm to read the data and display it properly.

Larger Machine Learning Datasets

In order to have a more effective machine learning model for extracting recipe content from webpages, it seems like it'd be necessary to create a much larger collection of training data. We currently only have 363 recipes in our collection, and dramatically increasing the size should produce better results.