

Chatbot Report

System Description

Brief overview of my chatbot:

Given a dataset of recipes gathered using a web crawler, my chatbot will ask for the user's name, and then prompt the user to describe or name the recipe that they want. The chatbot will then take the user input and find the recipe that most closely matches what the user wanted, if such a recipe exists in the dataset. It will remember what the user had asked for in the past, as well as the 5 most recent recipes that the user has encountered. On subsequent runs of the chatbot, the chatbot will take the data from its memory for that user and recommend a recipe based on the top 3 tags that the user most commonly uses. It will not recommend a recipe from the 5 previous recipes the user has searched (the 5 previous recipes does not include any recipe recommended by the chatbot).

One thing I want to note: if you manually modify the values in the `user_model`, it could result in some errors being thrown as the cursor is in an unexpected place, resulting in data being written in the wrong place. Therefore, do not manually modify the values in the `user_model` unless you need to.

More in depth overview/NLP concepts used:

First step – gathering data

I decided to use a website called based.cooking in order to gather recipe data. The home page has a list of the 50 most recently added recipes that I could use for my dataset. I utilized a web crawler to grab data from this website and properly place the data in a csv file. This data set was collected specifically for this project. It is not the same dataset that I had collected for my web crawler assignment.

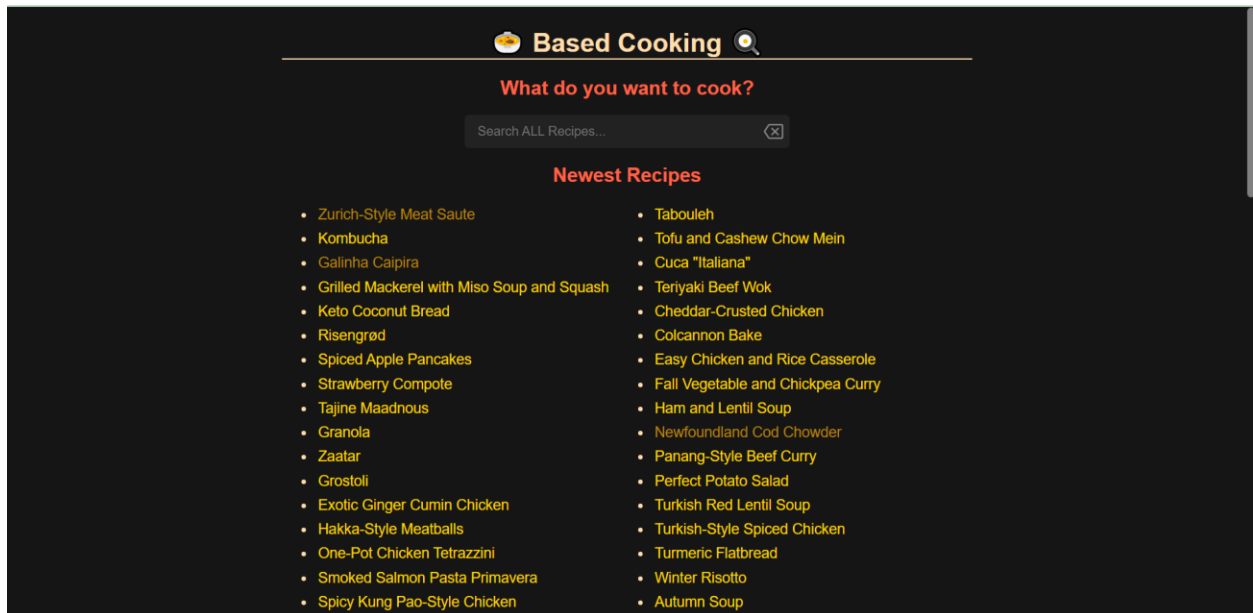


Figure 1: an image of the based.cooking website where I grabbed the data from

Based.cooking turns out to be a very good website to scrape data from, as it is a very simply designed website with no frills, which resulted in very clean raw data with little effort. After successfully getting the data into the csv, I went into the csv and manually cleaned anything that I couldn't clean before. I also wrote some tags for each recipe that I considered to be relevant, such as whether it was good for dinner or lunch, what important ingredients were in the recipe, and the total time (prep and cook) for the recipe in minutes.

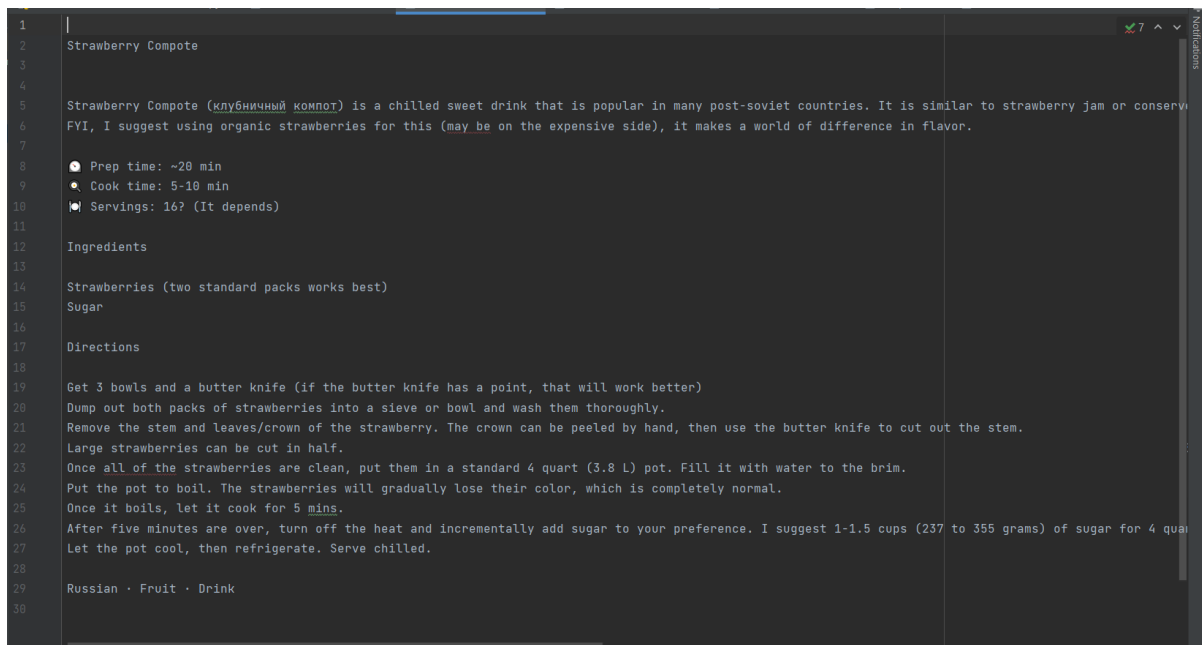
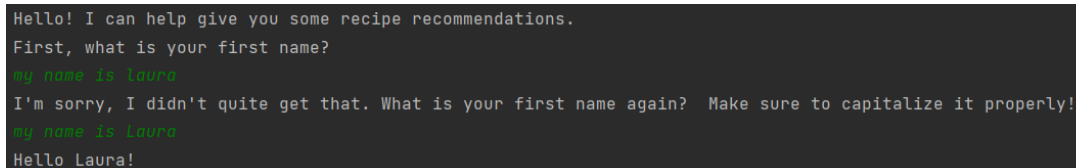


Figure 2: an image of the raw data I was able to scrape from the website

Second step – Greeting the User/Getting their Name

The first thing my chatbot will do is prompt the user for their first name. I needed to make this flexible, since the name could be anywhere in the string. I did some preprocessing on the user's input by first expanding out any contractions, since NLTK doesn't do a good job handling those. Then, using NLTK, I tokenized and pos tagged each word. Using the pos tags, I could search the input for anything that is a proper noun. If capitalized and used properly in the sentence, the name should be tagged as an NNP. This of course means that if the user does not capitalize their name that the chatbot won't recognize it. I get around this by simply asking the user for their name again and emphasizing that they should capitalize it properly. In addition, capitalizing other words other than the name can give incorrect results, but this is a tradeoff I intentionally took for reasons explained in the next paragraph. Therefore, in testing this, be sure to properly capitalize a name. If the user simply just types their name and nothing else, the pos tagger won't tag it as a proper noun. In this case, I check to see if the user input was a noun, and if it was then I choose that as the name.

I would like to note that I intentionally chose this way to do it as opposed to using NER systems because I noticed that those systems are rather inflexible when it comes to recognizing names. For instance, I had a NER system put in place before, but it had incorrectly mislabeled the name "Greg" as an organization rather than a person. With my system, while it may potentially accept some strings that are not names, it is more flexible in determining what the name is as opposed to the NER system I had previously investigated.



```
Hello! I can help give you some recipe recommendations.  
First, what is your first name?  
my name is laura  
I'm sorry, I didn't quite get that. What is your first name again? Make sure to capitalize it properly!  
my name is Laura  
Hello Laura!
```

Figure 3: an example of when the chatbot is unable to recognize the name because of improper capitalization.

Third step – Recommending a Recipe

The user now has an opportunity to tell the chatbot what kind of recipe it wants. Since this prompt is open-ended, I needed to preprocess the text to only get the relevant data. Examples of relevant data would be ingredient names, time that recipe should take, cuisine, and title of the recipe.

I word tokenized the user input, then made the tokens lowercase, removed stopwords, and removed punctuation (but did not remove numbers). I then lemmatized my tokens, since my dataset included tags that were in lemmatized form. Next, I had created a list of words that I often encountered in my testing that I considered to be useless and filtered those words out of my tokens. This of course would not be able to completely remove every non-relevant word, but it does a pretty good job of getting rid of most of the problematic words. Finally, in the case

where the user uses hours to describe the time instead of minutes, my program will convert that to minutes.

```
# list of some common words that are for certain not important for me to keep
non_important_words = ['please', 'give', 'recipe', 'want', 'need', 'have', 'keep', 'make', 'would', 'something',
                       'prefer', 'take', 'contain', 'contains', 'could', 'recommend', 'made', 'good', 'provide', 'around']
```

Figure 4: examples of non-important words that could occur in the user input

To determine which recipe to pick, I utilized a scoring system. Each recipe is assigned a score based on how closely the recipe matched the preprocessed user input. I gave 2 points to a recipe if a token from the user input matched the recipe title, since if the user is trying to look up a recipe by name, then that should be prioritized. 1 point is given if the user input matched one of the tags of the recipe, and 1 point is given if the time that the user had inputted is within 5 minutes of the time in the “Tags” section. Once scores were determined, I sorted the scores in descending order. If there was an outlier (1 recipe got more points than any other) then that is the recipe that will be shown. If the highest score is a 0, then none of the user tokens matched the dataset, and a random recipe will be shown to the user. If there are multiple recipes with the highest score, then I will collect all the ones with the highest scores and randomly pick one of them to show to the user. In the user model, all the preprocessed tokens said by the user are stored in a dict. In addition, the id of the recipe shown to the user in this step will also be stored in the user model.

Fourth step – Recommend Existing Users a Recipe

Using the user tokens and the 5 most recent recipes collected in step 3, my chatbot will recommend a recipe that is likely to fit the user’s interests. Recommendation of recipes can only happen if the user already exists in the user base. The user tokens are stored in a dict which keeps track of each token that has been used in scoring. However, it’s not a guarantee that all of the tokens stored are useful to scoring. I am using term frequency to determine what are the most relevant tokens for the user. I then do the same scoring process as step 3 to find a recipe to recommend to the user. I don’t want to recommend a recipe that the user has searched up recently, so the chatbot will use the 5 most recent recipes that were searched up by the user. There are situations where the chatbot won’t be able to recommend a recipe, which are shown in the sample dialogs.

Summary of NLP/class concepts used for the chatbot

Web crawler – collecting of data

Parse user input (NLTK tokenization, stopwords) – used any time the chatbot receives user input

POS tagging – used to determine the name of the user.

Lemmatization – for preprocessing some user input; used on recipe title in scoring to determine if a user tag matches a part of the recipe title.

Term frequency – utilized in the user model dictionary to determine the key words used the most by the user; used in the scoring system to match user input to a recipe.

Pandas – utilized for updating csv files

Diagram of Logic

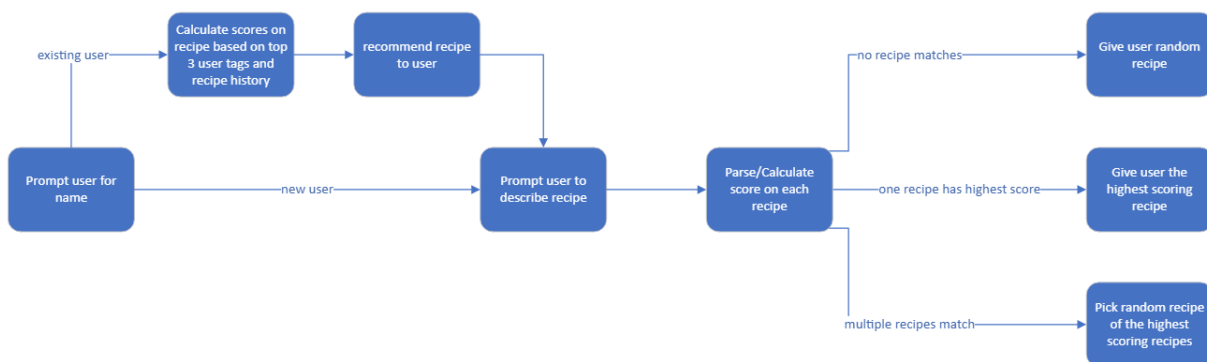


Figure 5: Logic diagram of my chatbot

Sample Dialog Interactions

White text is chatbot, green text is user input

Dialog One

User Sue is created, and a recipe is looked up. There is only one Russian recipe in the dataset, so the chatbot is unable to recommend a new recipe that the user hasn't seen.

```
Hello! I can help give you some recipe recommendations.
First, what is your first name?
my name is sue
I'm sorry, I didn't quite get that. What is your first name again? Make sure
to capitalize it properly!
my name is Sue
Hello Sue!
What kind of recipe are you looking for?
give me a russian recipe
```

one recipe found

Here is a recipe that you may like based on what you wanted!

Strawberry Compote

Prep time: ~20 min

Cook time: 5-10 min

Servings: 16

Ingredients:

Strawberries (two standard packs works best)

Sugar

Directions:

Get 3 bowls and a butter knife (if the butter knife has a point, that will work better)

Dump out both packs of strawberries into a sieve or bowl and wash them thoroughly.

Remove the stem and leaves/crown of the strawberry. The crown can be peeled by hand, then use the butter knife to cut out the stem.

Large strawberries can be cut in half.

Once all of the strawberries are clean, put them in a standard 4 quart (3.8 L) pot. Fill it with water to the brim.

Put the pot to boil. The strawberries will gradually lose their color, which is completely normal.

Once it boils, let it cook for 5 mins.

After five minutes are over, turn off the heat and incrementally add sugar to your preference. I suggest 1-1.5 cups (237 to 355 grams) of sugar for 4 quarts (3.8 L) of compote. Taste between helpings of sugar to make sure its not oversweetened.

Let the pot cool, then refrigerate. Serve chilled.

Hello! I can help give you some recipe recommendations.

First, what is your first name?

Sue

Welcome again Sue!

Here is a recipe you may like based on what you have looked at before:

your top 3 search queries are: ['russian']

I couldn't find a new recipe that you haven't seen yet. Keep searching!

What kind of recipe are you looking for?

Dialog Two

User Nate is created, and a more sophisticated search is done to find a recipe. Multiple recipes are found, and a recipe is chosen that accommodates some of the tags. On the next run, a recipe is recommended to Nate.

Hello! I can help give you some recipe recommendations.

First, what is your first name?

you can call me Nate

Hello Nate!

What kind of recipe are you looking for?

can you give me a recipe that contains onions, is good for dinner, and can be made in 40 minutes?

multiple recipes found!

Here is a recipe that you may like based on what you wanted!

Shrimp and Chicken Jambalaya

Prep time: 20 min

Cook time: 40 min

Servings: None

Ingredients:

1/4 tsp Cayenne

3/4 tsp White Pepper

1 tsp Kosher Salt

1/4 tsp Thyme

1/2 tsp Rubbed Sage

1/4 tsp Dried Basil

1/2 tsp Black Pepper

1 Tbsp Unsalted Butter

1/2 cup Diced Andouille (or Chirozo sausage)

1/2 cup Diced Onion

1/2 cup Diced Bell Pepper

1/2 cup Diced Celery

1/2 cup Diced Fresh Tomatoes

1/2 cup Tomato Sauce

3/4 cup Enriched Long Grain Rice

1-1/4 cup Chicken Stock

1 Tbsp Worcestershire Sauce

2 Tbsp Minced Fresh Garlic

1/2 cup Diced Chicken (cooked or raw)

1-1/2 cup Medium Shrimp

1 Tbsp Finely Chopped Italian Parsley

3 Tbsp Finely Sliced Green Onions

Directions:

Preheat the oven to 350 degrees F. Mix together the seasoning mix (herbs, salt, pepper).

Mix together the Holy Trinity (onion, celery, bell pepper).

In a Cast Iron Dutch Oven, melt the butter over medium heat, add the Andouille and cook until it starts to brown.

Add 1/2 of the Holy Trinity, cook until the vegetables are tender.

Add the diced tomatoes and cook for one minute. Add the tomato sauce and cook for another minute.

Add the rice and cook for 2 minutes, stirring constantly.

Add the stock, remaining Holy Trinity, seasoning mix, Worcestershire, and the garlic.

Taste the broth for seasoning, particularly salt.

Add the chicken, stir well and put the pot in the preheated oven. Bake uncovered for 25 minutes.

Stir in the raw shrimp, parsley, and green onions - place back into oven for another 10 minutes (or until the shrimp is cooked through).

Hello! I can help give you some recipe recommendations.

First, what is your first name?

nice to see you again, my name is Nate!

Welcome again Nate!

Here is a recipe you may like based on what you have looked at before:

your top 3 search queries are: ['onion', 'dinner', '40']

Easy Chicken and Rice Casserole

Prep time: 5 min

Cook time: 40 min

Servings: 4

Ingredients:

1/2 cup Onion, chopped

1/2 cup Celery, chopped

2 Tbsp Butter or margarine

2 cups Chicken, cooked and chopped

1 3/4 cups Hot water

2/3 cup Long Grain White Rice

1 can (10 oz) Mushrooms with liquid

1 cup Frozen Peas and Carrots

1 tsp Dried Thyme and Rosemary

Directions:

In large fry pan, cook onion and celery in butter until soft.

Stir in remaining ingredients.

Bake in covered 8 cup baking dish in preheated (350F) oven for about 30 minutes or until rice is cooked.

What kind of recipe are you looking for?

Dialog Three

User Natalie is created, but Natalie searches for the tag “korean,” which does not exist in my dataset. Natalie is shown a random recipe instead, which is added to her user data. Because the tag was unable to get a result, I did not add it to Natalie’s top terms. The recommendation system is therefore unable to recommend anything because of her tags.

Hello! I can help give you some recipe recommendations.

First, what is your first name?

I'm Natalie

Hello Natalie!

What kind of recipe are you looking for?

do you have any korean recipes I can make?

no matches found

Sorry, I was not able to find a recipe that has what you want.

Here is a random recipe you may like!

Colcannon Bake

Prep time: 20 min

Cook time: 70 min

Servings: 6

Ingredients:

3 Potatoes, peeled and quartered

7 Tbsp Butter, cut into small chunks

1/2 cup Sour Cream

1 Egg
1 Tbsp Milk
3 cup Cabbage, shredded
2 Leeks, chopped
1 small Onion, chopped
2 cubes Chicken Bouillon
1/2 cup Cheddar Cheese, shredded

Directions:

Place the potatoes into a large pot and cover with salted water. Bring to a boil, then reduce heat to medium-low, cover and simmer until tender, about 20 minutes. Drain and allow to steam dry for a minute or two. Season the potatoes with salt, and mash with 6 Tbsp of the butter, the sour cream, egg, and milk.

Preheat oven to 350F. Grease a 2-quart casserole dish.

Heat the remaining butter in a skillet over medium heat. Stir in the cabbage, leeks, and onion until the cabbage is tender and the onion is translucent (10 mins).

Crush the bouillon cubes into the cabbage mixture, and stir to blend and dissolve the cubes. Stir the cabbage mixture into the potato mixture until thoroughly mixed, and spoon into the prepared casserole dish.

Bake in the preheated oven for 40 minutes, top with cheddar cheese, and return to oven until the cheese melts (10 mins).

Hello! I can help give you some recipe recommendations.

First, what is your first name?

my name is Natalie

Welcome again Natalie!

Here is a recipe you may like based on what you have looked at before:

your top 3 search queries are: []

I can't seem to find a different recipe based on your search terms. Try looking up some different recipes!

What kind of recipe are you looking for?

Dialog Four

An existing user Frank is used with the following user model:

```
Frank,"{'egg': 2, '30': 3, 'minute': 5, 'salad': 1, 'dinner': 3, 'breakfast': 2, 'onion': 1, 'lunch': 2, 'cilantro': 1, '20': 1, 'easy': 1, '60': 1, 'minutes': 1, '10': 1}","['10', '6', '46', '8', '7']"
```

Frank gets a personalized recommendation based on his preferences. He then looks up the word ‘pancake’ which matches the title of a recipe.

your top 3 search queries are: []

Hello! I can help give you some recipe recommendations.

First, what is your first name?

Frank is my name

Welcome again Frank!

Here is a recipe you may like based on what you have looked at before:

your top 3 search queries are: ['minute', '30', 'dinner']

Hakka-Style Meatballs

Prep time: 10 min

Cook time: 20 min

Servings: 3

Ingredients:

1/2 lb Ground Beef
1/4 cup Panko Breadcrumbs
2 Tbsp Soy Sauce, divided
4 tsp Chili Garlic Sauce
1/2 oz Ginger, minced
2 medium Green Bell Peppers
1 small Onion, chopped
3/4 cup Basmati Rice
2 oz Green Peas
1/2 cup Cilantro, loosely packed
1 Tbsp Moroccan Spice Blend, divided
4 Tbsp Sweet Chili Sauce
Oil

Directions:

Preheat oven to 450F. Add 1 1/4 cup water and 1/4 tsp salt to a medium pot. Cover and bring to a boil over a medium-high heat. Add rice and reduce heat to low. Cover and cook until rice is tender and liquid is absorbed (12-14 mins). Add peas halfway through cooking. When rice is done, remove pot from heat, and leave covered.

While rice cooks, combine beef, panko, half the ginger, a quarter of the soy sauce and a third of the Moroccan spice blend. Roll mixture into 9 equal-sized meatballs. Arrange meatballs on a parchment-lined baking sheet. Bake on the middle rack in the oven until golden-brown and cooked through (8-10 mins).

Meanwhile, core and cut the green peppers into 1/2" pieces and roughly chop cilantro. Stir together sweet chili sauce, remaining soy sauce, 1/4 cup water and chili garlic sauce in a small bowl.

Heat a large non-stick pan over medium-high heat. Add 1 Tbsp oil, then onion and peppers. Cook, stirring occasionally, until slightly softened (2-3 mins). Add remaining ginger and remaining Moroccan spice blend. Cook, stirring constantly, until fragrant (30 sec).

Add meatballs and the stir-fry sauce to the pan with veggies. Cook, stirring gently, until sauce thickens slightly and coats meatballs (1 min). Season with salt and pepper, to taste.

Add half the cilantro to rice, then fluff with a fork. Divide rice between plates, then top with meatballs, veggies, and any sauce in the pan. Sprinkle remaining cilantro over top.

What kind of recipe are you looking for?

get me a pancake recipe

one recipe found

Here is a recipe that you may like based on what you wanted!

Spiced Apple Pancakes

Prep time: 15 min

Cook time: 10 min

Servings: 8 pancakes

Ingredients:

1 cup flour

1 large apple, peeled

1 tbsp sugar

1 tbsp cinnamon (or more to taste)

1/2 tsp baking powder

1 tsp vanilla extract

2/3 cup milk

1 egg

Directions:

In a medium bowl, combine flour, sugar, cinnamon, baking powder, and 1 pinch of salt.

Separate the egg yolk and egg white. Add the egg yolk to the dry ingredients.

Whisk the egg white until foamy.

Slice the peeled apple into thin strips. Alternatively, grate the apple.

Combine dry ingredients, apple, vanilla extract, egg white, and milk into a homogenous batter.

To a large pan, add 1 tbsp butter. Add the batter (about 3 tbsp per pancake) and cook for 2 min per side on medium-high heat.

Knowledge Base Appendix

Note: if something is underlined then I got it from based.cooking. Italicized means that its something that I added to the data.

Id – unique integer for each recipe.

Name – name of the recipe.

Prep time – time to prepare ingredients for the recipe.

Cook time – time to cook the prepared ingredients.

Servings – number of portions made from the recipe.

Ingredients – list of ingredients used for the recipe and the amounts used of each one.

Directions – list of instructions to create the recipe.

Tags – list of words that are relevant to the recipe, including ingredient names, total cook times (calculated by adding the prep time and cook time), whether the recipe is quick (I decided that a quick recipe is one that takes 30 minutes or less, with the exception of any recipe that was already described as “quick” on based.cooking), what time of day would be good for the recipe (breakfast, lunch, dinner), etc.

| id | Name | Prep Time | Cook Time | Servings | Ingredient: Directions: Tags | | | | | | | | | | | | | | | | | | | |
|----|---------------------------------|-----------|-----------|----------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Tajine Maadnouns | 20 min | 35 min | 12 | ['1 lb chick', 'Mix the s', 'tunisian', 'eggs', 'dinner', 'chicken', 'onion', 'potato', 'parsley', 'cheese', 'cumin', 'coriander', 'turmeric', 'paprika', 'cayenne', 'caraway', 'thyme', 'cinnamon', '55'] | | | | | | | | | | | | | | | | | | | |
| | Perfect Potato Salad | 15 min | 5 min | 8 | ['5 Potatos', 'Place pot', 'american', 'quick', 'fast', 'side', 'yogurt', 'mayonnaise', 'horseradish', 'radish', 'celery', 'green onion', 'parsley', 'dill', '20'] | | | | | | | | | | | | | | | | | | | |
| 3 | Strawberry Compote | ~20 min | 5-10 min | 16 | ['Strawber', 'Get 3 box', 'russian', 'fruit', 'drink', 'quick', 'fast', 'sugar', '30'] | | | | | | | | | | | | | | | | | | | |
| | Easy Chicken and Rice Casserole | 5 min | 40 min | 4 | ['1/2 cup C', 'In large fi', 'quick', 'dinner', 'lunch', 'onion', 'celery', 'butter', 'margarine', 'mushrooms', 'peas', 'carrots', 'thyme', 'rosemary', '45'] | | | | | | | | | | | | | | | | | | | |
| 5 | Turkish Red Lentil Soup | 10 min | 30 min | 6 | ['1 cup Rex', 'Reserve f', 'vegetarian', 'dinner', 'olive oil', 'paprika', 'red pepper flakes', 'onion', 'carrot', 'tomato paste', 'cumin', 'mint', 'thyme', 'oregano', 'vegetable broth', 'broth', '40'] | | | | | | | | | | | | | | | | | | | |
| | Hakka- | | | | | | | | | | | | | | | | | | | | | | | |

Here is an example of a specific item:

| Directions | Tags |
|--|--|
| ['Combine panko, cheese, and smoked paprika-garlic blend in a shallow dish. Preheat oven to 425F.\n', 'Pat chicken dry with paper towel. Carefully slice into the centre of each chicken breast lengthwise and parallel to the cutting board - leaving 1/2 inch intact on the other end. Open up chicken like a book, and season both sides with salt and pepper.\n', 'Coat one side of each chicken breast with mayo. Firmly press mayo-coated side into panko mixture to adhere, one at a time.\n', 'Transfer chicken to a parchment-lined baking sheet, coated-side up. Bake chicken on the middle oven rack until cooked through (18-20 mins).\n', '\n'] | ['quick', 'fast', 'dinner', 'lunch', 'mayonnaise', 'panko breadcrumbs', 'cheese', 'smoked paprika-garlic blend', '30'] |

Name – first name of the user

Recipe_history – list of the id numbers of the up to 5 most recent recipes that the user has looked up, not including recipes that were recommended to them by the chatbot.

Chatbot Evaluations

My chatbot has a specific purpose to recommend and find recipes that users would want to see. It's very good at carrying out that specific purpose in the way that I had wanted. All the user needs to do is write out what they want, and at the end they will be provided with some recipe that matches at least some of what they wanted, or lets the user know that what they want doesn't exist in the dataset. In that way, the purpose of my chatbot is easy to understand. Simply say what you want and the chatbot does its best to give you what you want. Finally, the dataset that I collected contained a wide range of recipes with different qualities despite its small size.

There are a few things that I think could be improved on this chatbot. For one, the user can't specify to the chatbot things that they would not want in their recipe. For instance, if the user asks for no chicken, then the chatbot will not know how to handle that and will just give recipes with chicken. Also, once the chatbot is closed, there is no way for the user to access the recipes they had searched for unless they search using the exact same keywords. Finally, while I had mentioned that despite its size my dataset had a unique spread of recipes, it didn't contain a lot of recipes that I feel that people would search up often. I would need to either make my dataset larger or find a way to scrape for popular recipes, or both.

Overall, while my chatbot is lacking in some ways, I'm satisfied with how it turned out. I set out and achieved what I had in mind for my chatbot.