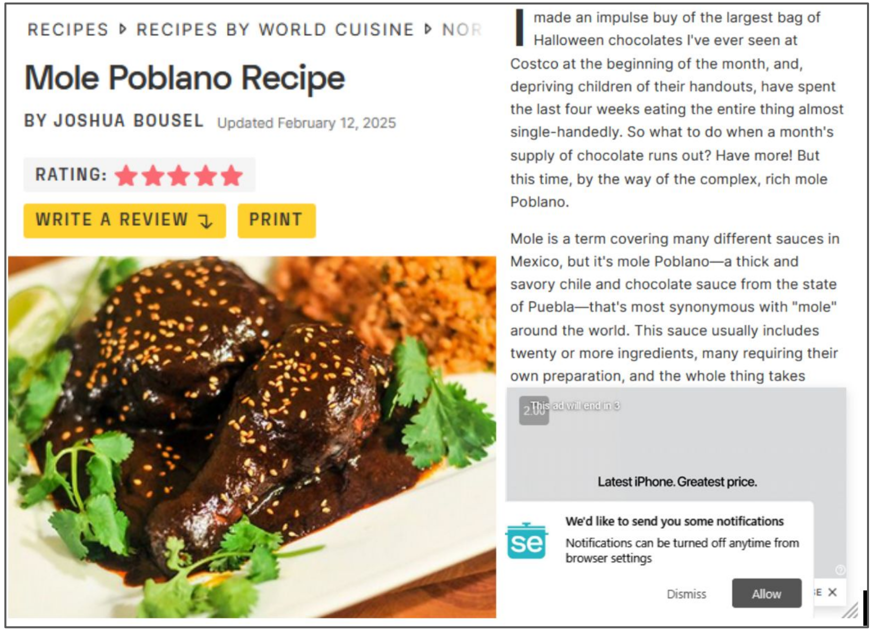
Advanced Recipe Database

Kyle Watson
Luke Unterman
Paula Gearon

Problem: Most Recipe Websites Suck

- Highest-ranked web pages retrieved by a search engine (e.g., Google) are:
  - Heavily bloated with extraneous information (see Figure 1)
  - Polluted with unskippable advertisements that detract from overall user experience
- They also frequently lack support for
  - Unit conversions (e.g., ounce => gram)
  - Ingredient substitutions
- Band-aid solutions do not solve root issue:
  - Website design itself is flawed
  - Enter: RRDBMS (Recipe Relational Database Management System)


RECIPES ▸ RECIPES BY WORLD CUISINE ▸ NOR

Mole Poblano Recipe

BY JOSHUA BOUSEL Updated February 12, 2025

RATING: ★★★★★

WRITE A REVIEW ↴ PRINT

I made an impulse buy of the largest bag of Halloween chocolates I've ever seen at Costco at the beginning of the month, and, depriving children of their handouts, have spent the last four weeks eating the entire thing almost single-handedly. So what to do when a month's supply of chocolate runs out? Have more! But this time, by the way of the complex, rich mole Poblano.

Mole is a term covering many different sauces in Mexico, but it's mole Poblano—a thick and savory chile and chocolate sauce from the state of Puebla—that's most synonymous with "mole" around the world. This sauce usually includes twenty or more ingredients, many requiring their own preparation, and the whole thing takes

This ad will end in 3

Latest iPhone. Greatest price.

We'd like to send you some notifications

Notifications can be turned off anytime from browser settings

Dismiss Allow

Figure 1. Mole Poblano recipe viewed on iPhone 13 Pro Max.
- Left: recipe with image thumbnail.
- Right: image of 316 word preamble (obscured by video ad) preceding actual recipe about author's personal journey in creating this recipe.

Proposed Design: Recipe RDBMS

- Advanced recipe database with searchable, interlinked, and flexible recipe structure
  - Don’t have a particular recipe ingredient? Replace it with a valid substitution
  - Prefer the metric system? Each unit has a list of potential conversions.
- Image similarity search:
  - Given any input image, find recipes with similar images
- Text similarity search:
  - Given any input sentence, find recipes with similar descriptions/steps/titles
- Simplistic, easily navigable website design

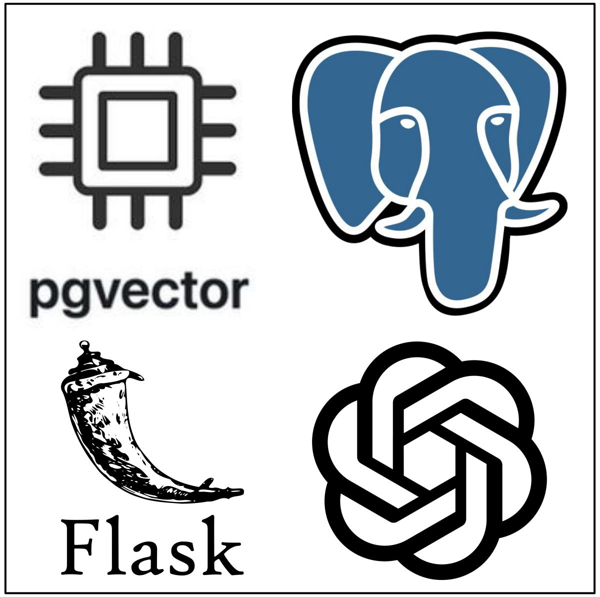


Figure 2. Recipe RDBMS tech-stack

- Using PostgreSQL database to store our data, with pgvector extension to store and index embeddings. Image + text embeddings created using OpenAI CLIP model and MPNet, respectively. We use Flask to host our website.

Principle Entities

The principle entities are:

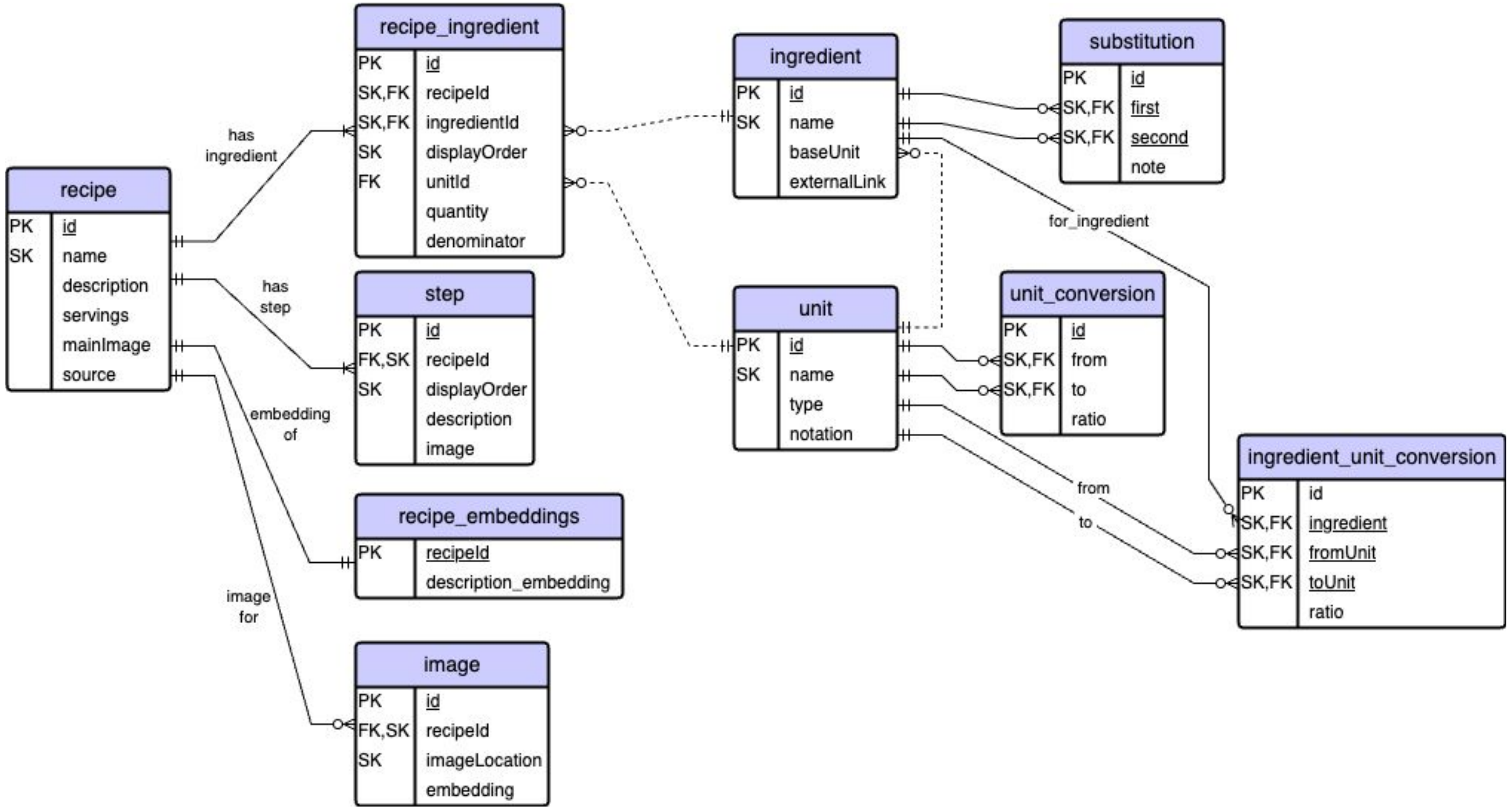
- Recipe
- Ingredients
- Units
- Images

Supporting Entities

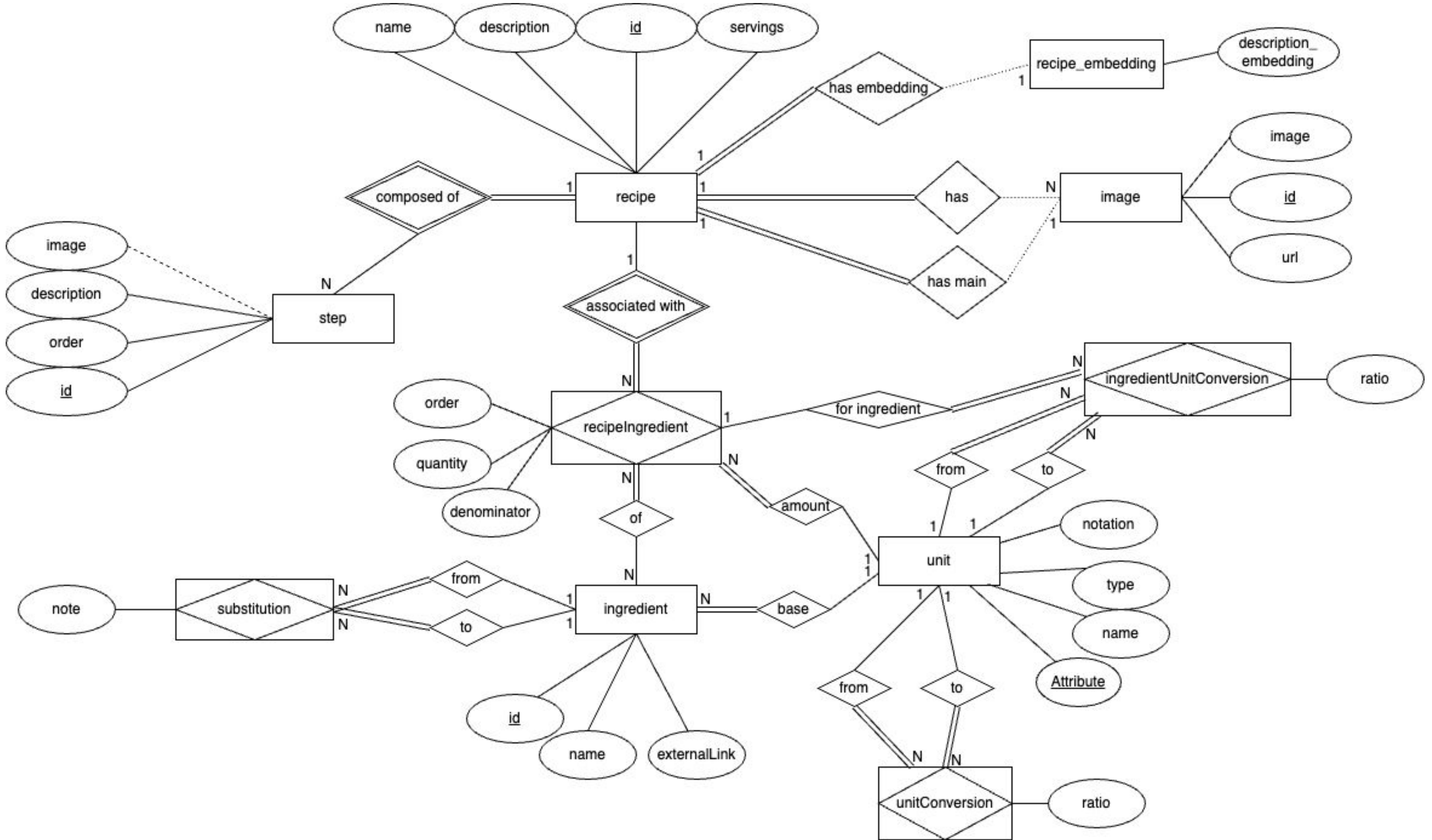
Other supporting entities are:

- Recipe Ingredients
- Unit Conversions
- Ingredient Unit Conversions (e.g. 1 cup flour ≡ 125g flour)
- Unit Types
- Recipe Steps
- Embedding Vectors

Crow’s Foot ER Diagram



Chen ER Diagram



Data Acquisition + Preprocessing

- Using Python script, retrieve all SeriousEats URLs from XML sitemap ending with “recipe”
    - For each website, scrape JSON-LD object containing structured recipe information
- Preprocess dumped JSON object to only contain required fields (recipe steps, ingredients, description, etc.)
    - For each recipe ingredient:
        - Use Python script to connect to Google Gemini 1.5 Flash LLM
        - Prompt LLM to split ingredient into quantity, unit, and food
            - “ 1 cup broccoli florets” => {Q: 1, U: Cup, F: broccoli florets}
- Iteratively create SQL script to insert values into our database
- **Note:** this was probably more trouble than it was worth

Image Search

- Using CLIP model to convert recipe images to embedding vectors.

“CLIP is a is a multimodal vision and language model motivated by overcoming the fixed number of object categories when training a computer vision model.”

- Stored with datatype `vector(512)` provided by *pgvector*

```
SELECT image.id, recipeId, imageLocation,
       (embedding <=> %s::vector(512)) AS cosine_distance,
       recipe.name as recipe_name
FROM image
JOIN recipe on image.recipeid = recipe.id
WHERE embedding IS NOT NULL
ORDER BY cosine_distance
LIMIT 5;
```

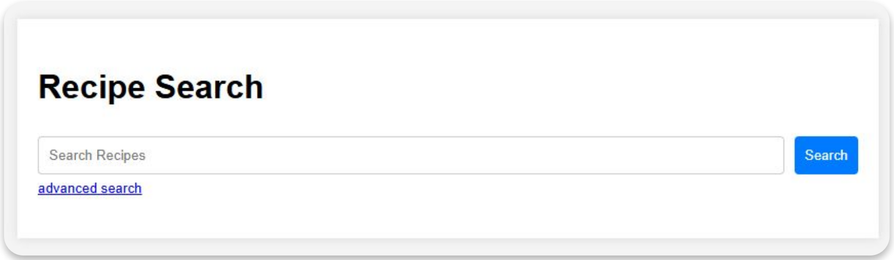
OpenAI CLIP

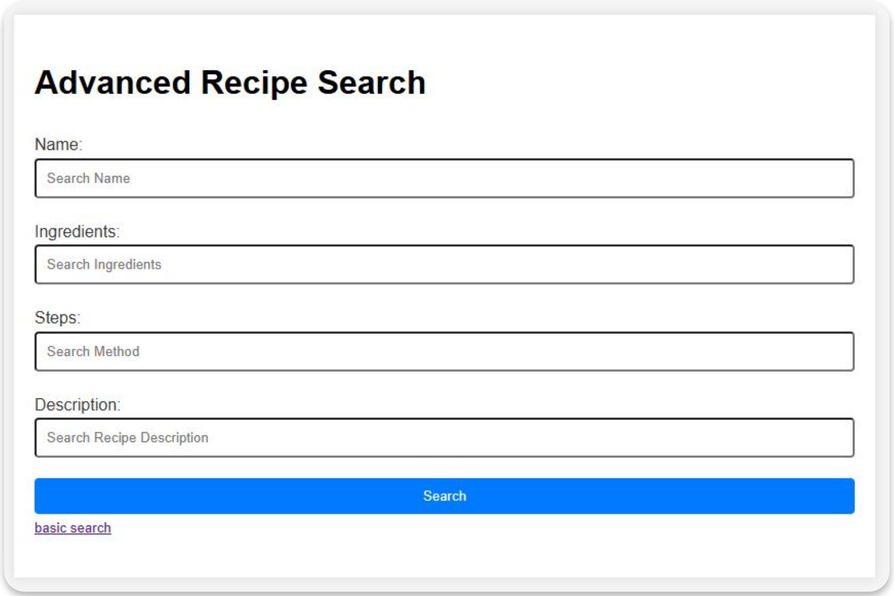
CONNECTING TEXT AND IMAGES

Fuzzy and Advanced Text Search

- Fuzzy search
  - Similar to image search
  - Combined name, description, ingredients, and steps into a vector “embedding”
  - Compare query embedding to saved embeddings
  - Return sorted results based on embedding similarity

- Advanced search
  - Matches exact names, descriptions, ingredients, and steps





Text Search Implementation

- “Embed” text descriptions using a BERT-based model
  - SBERT
- Stored with datatype: `vector(768)`
- Datatype provided by *pgvector* extension to Postgres

```
SELECT id, name, mainimage, description from recipe
INNER JOIN recipe_embeddings on
recipe.id=recipe_embeddings.recipeid
ORDER BY description_embedding <#> %s::vector ASC LIMIT 10"
```

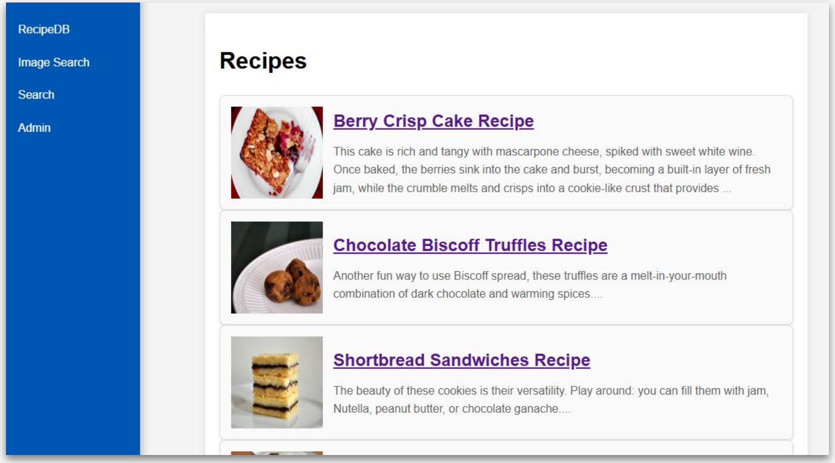
Flask

A simple Python framework for building Web applications and services.

- Static Files
  - HTML, CSS, JS, image
- Generated pages
  - Jinja Templates
- Web Services

Website Design and Styling

- /index page lists several recipes
- Navbar on every page to link to main pages
- Advanced search inside the text search tab
- Navbar moves to the top of the screen when resolution is too narrow
- Recipe ‘cards’ summarize and link to recipe page
  - Scale when hovered

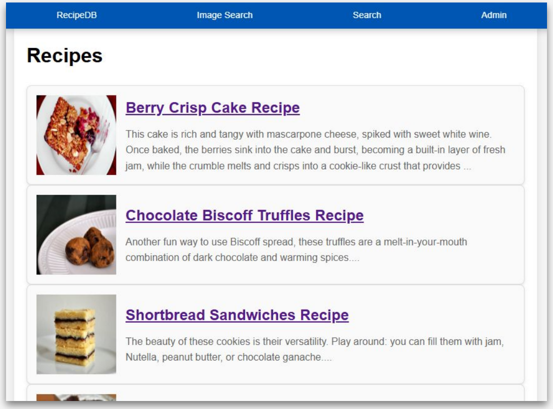

RecipeDB
Image Search
Search
Admin

Recipes

Berry Crisp Cake Recipe
This cake is rich and tangy with mascarpone cheese, spiked with sweet white wine. Once baked, the berries sink into the cake and burst, becoming a built-in layer of fresh jam, while the crumble melts and crisps into a cookie-like crust that provides ...

Chocolate Biscoff Truffles Recipe
Another fun way to use Biscoff spread, these truffles are a melt-in-your-mouth combination of dark chocolate and warming spices....

Shortbread Sandwiches Recipe
The beauty of these cookies is their versatility. Play around: you can fill them with jam, Nutella, peanut butter, or chocolate ganache....


RecipeDB Image Search Search Admin

Recipes

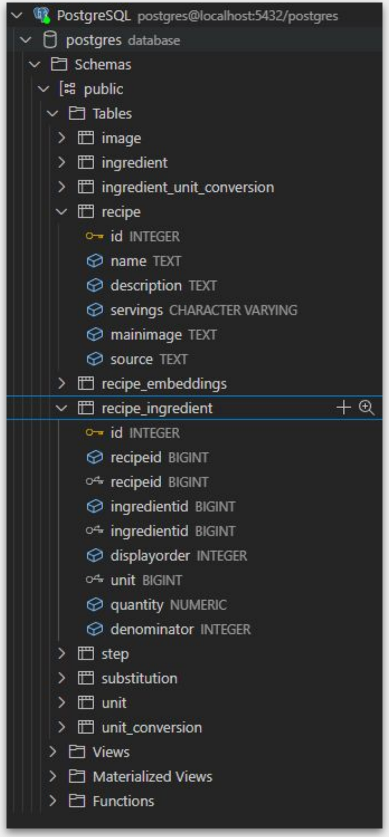
Berry Crisp Cake Recipe
This cake is rich and tangy with mascarpone cheese, spiked with sweet white wine. Once baked, the berries sink into the cake and burst, becoming a built-in layer of fresh jam, while the crumble melts and crisps into a cookie-like crust that provides ...

Chocolate Biscoff Truffles Recipe
Another fun way to use Biscoff spread, these truffles are a melt-in-your-mouth combination of dark chocolate and warming spices....

Shortbread Sandwiches Recipe
The beauty of these cookies is their versatility. Play around: you can fill them with jam, Nutella, peanut butter, or chocolate ganache....

DEMO

SQL Tools - VSCode Extension



```
PostgreSQL postgres@localhost:5432/postgres
  postgres database
    Schemas
      public
        Tables
          image
          ingredient
          ingredient_unit_conversion
          recipe
            id INTEGER
            name TEXT
            description TEXT
            servings CHARACTER VARYING
            mainimage TEXT
            source TEXT
          recipe_embeddings
          recipe_ingredient
            id INTEGER
            recipeid BIGINT
            recipeid BIGINT
            ingredientid BIGINT
            ingredientid BIGINT
            displayorder INTEGER
            unit BIGINT
            quantity NUMERIC
            denominator INTEGER
          step
          substitution
          unit
          unit_conversion
        Views
        Materialized Views
        Functions
```

- Connect to databases (not just PostgreSQL)
- View tables and properties
- Useful to have next to code and when crafting SQL queries