

Recuperación de Información y Recomendaciones en la Web

Curso 2024

Informe Final

Grupo 9

Docente: Libertad Tansini

Índice

- 1. Introducción - Problema a abordar**
- 2. Enfoque de solución**
- 3. Diseño - Arquitectura del sistema**
- 4. Implementación.**
- 5. Evaluación y resultado**
- 6. Conclusiones**
- 7. Trabajo Futuro**

Introducción - Problema a abordar

Un problema usual con el que todos nos encontramos es que a veces uno no tiene ideas para cocinar con los ingredientes que tiene en el hogar, por lo que poder encontrar recetas que se adapten a los ingredientes que se tienen parece un problema interesante.

Se planea realizar una aplicación que simplifique este proceso, permitiendo ingresar los ingredientes que se tienen y recibir una lista de posibles recetas que se preparen con los mismos, optimizando el tiempo dedicado a la planificación de comidas y fomentando la utilización efectiva de los recursos disponibles en casa.

Otra adición importante es la posibilidad de agregar restricciones alimentarias y preferencias dietéticas, por lo que se agregan filtros que resuelvan esta parte del problema, incluyendo opciones vegetarianas, sin lácteos, sin gluten y más.

Enfoque de solución

Desde un punto de vista técnico, se implementó una aplicación web full stack con Ruby on Rails, específicamente con Ruby 3.3.1 y Rails 7.1.3, además una base de datos postgres 16.3, un servicio de Elasticsearch 8.13.4, todo esto orquestado mediante docker para unificar ambientes de desarrollo.

Sobre la recuperación de las recetas se barajaron dos opciones, utilizar scraping para leerlas desde sitios de recetas de internet o consumirlas desde APIs, luego de investigación se optó por la segunda opción. Fueron consideradas varias APIs (Spoonacular, TheMealDB, Tasty, Edamam, etc...) aunque luego de varias pruebas se terminó utilizando Spoonacular y TheMealDB, ya que fueron las únicas que brindaban las instrucciones para las recetas en su versión gratis.

Yendo más en específico, fueron consumidas todas las recetas de TheMealDB (alrededor de 350) y una cantidad similar desde Spoonacular de manera aleatoria, logrando una cantidad total del entorno de 750 recetas. Luego de consumidas, las recetas son almacenadas en nuestra base de datos, en una tabla de recetas con un esquema construido para ser un subconjunto de la información que traen las dos APIs seleccionadas.

Una vez las recetas llegan a la base de la aplicación, se utiliza Elasticsearch, un motor de búsqueda distribuido que permite procesar grandes volúmenes de datos en tiempo real, que implementa índices y tablas de posting para permitir búsqueda sobre el texto de las recetas. Elasticsearch implementa una arquitectura de nodos que puede escalar horizontalmente, para este proyecto realizado solo se utilizó un único nodo, que almacena los índices y así poder buscar la información que el usuario desea para posteriormente mostrarla en la aplicación web.

Arquitectura del sistema

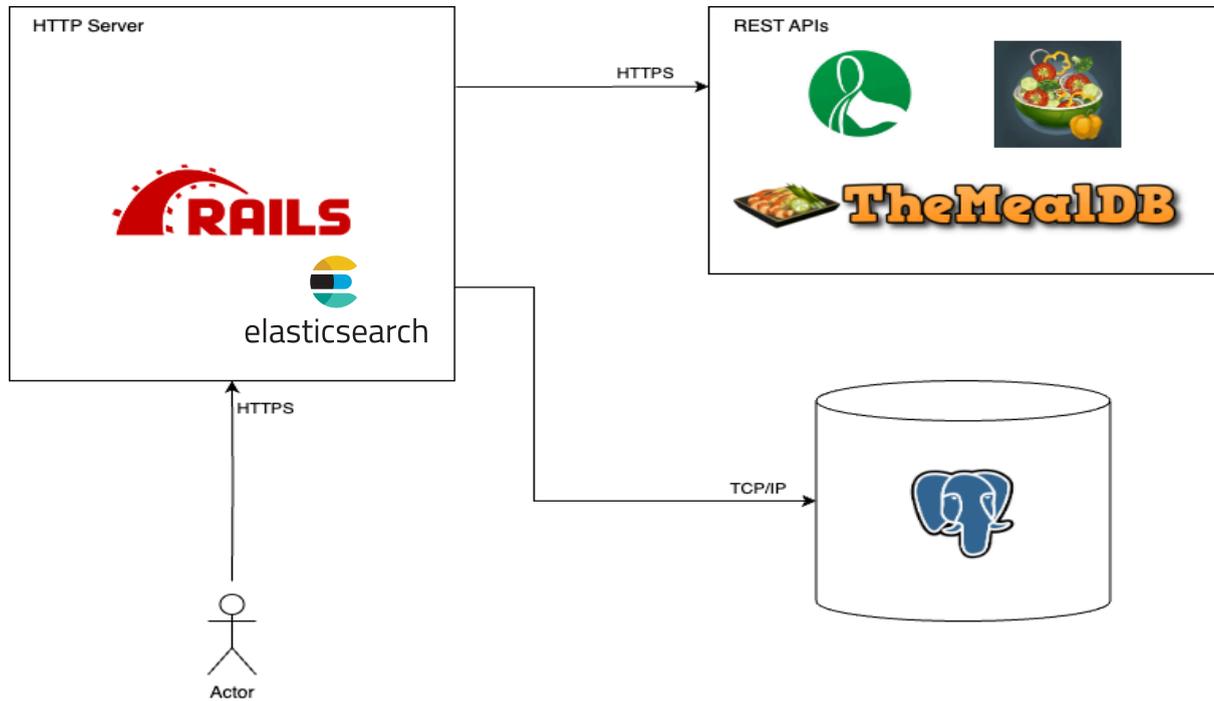


Figura 1: Arquitectura del sistema.

La arquitectura del programa ya fue explicada brevemente en la sección anterior. Sin embargo, es interesante destacar que se utilizó un ambiente de desarrollo dockerizado para que todos los integrantes del grupo tuviéramos la misma versión de cada tecnología y así no perder tiempo en instalaciones de tecnologías en distintos entornos de desarrollo ya sea windows o linux, con solo un par de comandos ya era posible implementar en un entorno de desarrollo local sin ningún problema.

Problemas encontrados

Inicialmente, se había elegido un proyecto de recomendación de playlists integrando spotify con un repositorio de información de música, sin embargo se encontraron varios problemas a la hora de lograr la integración. En específico, nos pareció que el problema pensado se podía reducir a el consumo de una sola API, por lo que luego de conversar con la docente, se optó por cambiar de proyecto, decantándose por el del recetario.

También ocurrió que varias de las APIs que se planeaban usar eran pagas o fuertemente limitadas en su versión gratis, lo más común es que en su versión gratis las recetas no tengan instrucciones o haya una fuerte limitación sobre la cantidad de API calls que podíamos hacer, lo que dificulta mucho nuestro proceso de ingestión de datos. Esto terminó implicando que tuvimos que descartar varias APIs.

Evaluación y resultado

A continuación se presentan algunas imágenes de la aplicación web.

En la Figura 2 se puede ver la página principal del buscador, es la primera página a la que el usuario es dirigido cuando entra a la web y donde el mismo puede ingresar un ingrediente para buscar una receta que lo contenga. Además puede marcar alguna de las opciones debajo del buscador en caso de tener alguna restricción o preferencia alimenticia.

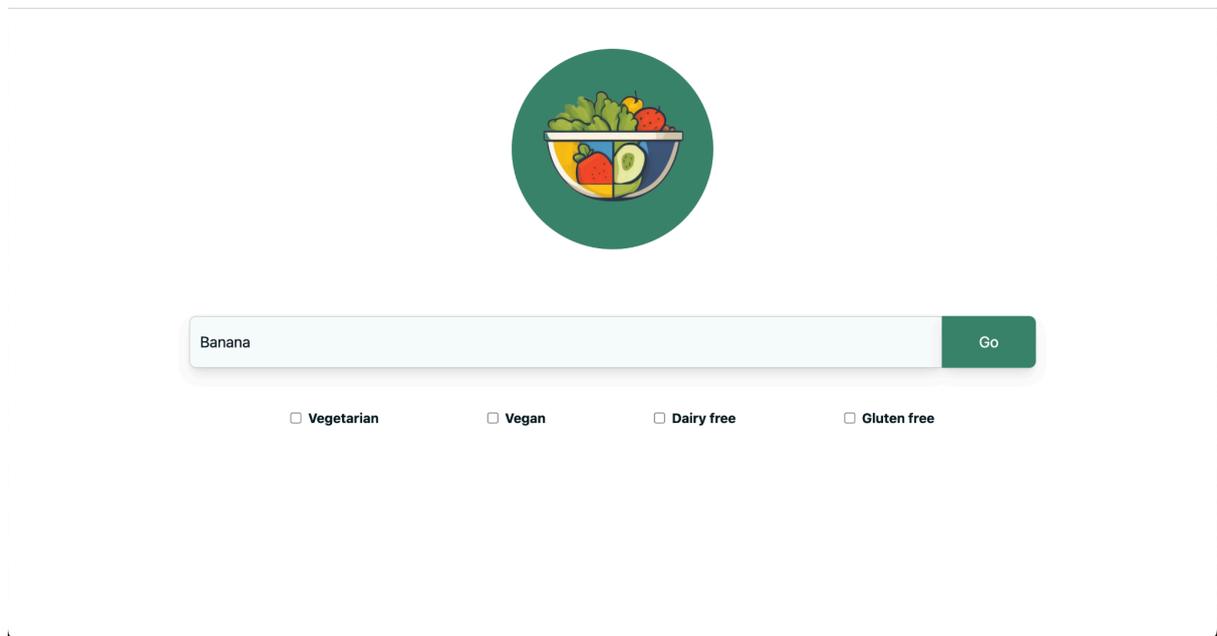


Figura 2: Página principal de la aplicación.

Cuando el usuario procede con la búsqueda, se listan los resultados con recetas que contengan el ingrediente que ingresó y con las restricciones alimenticias ingresadas. En la Figura 3 se puede ver la lista de recetas para el ingrediente “Banana” y en este caso sin ninguna restricción alimenticia

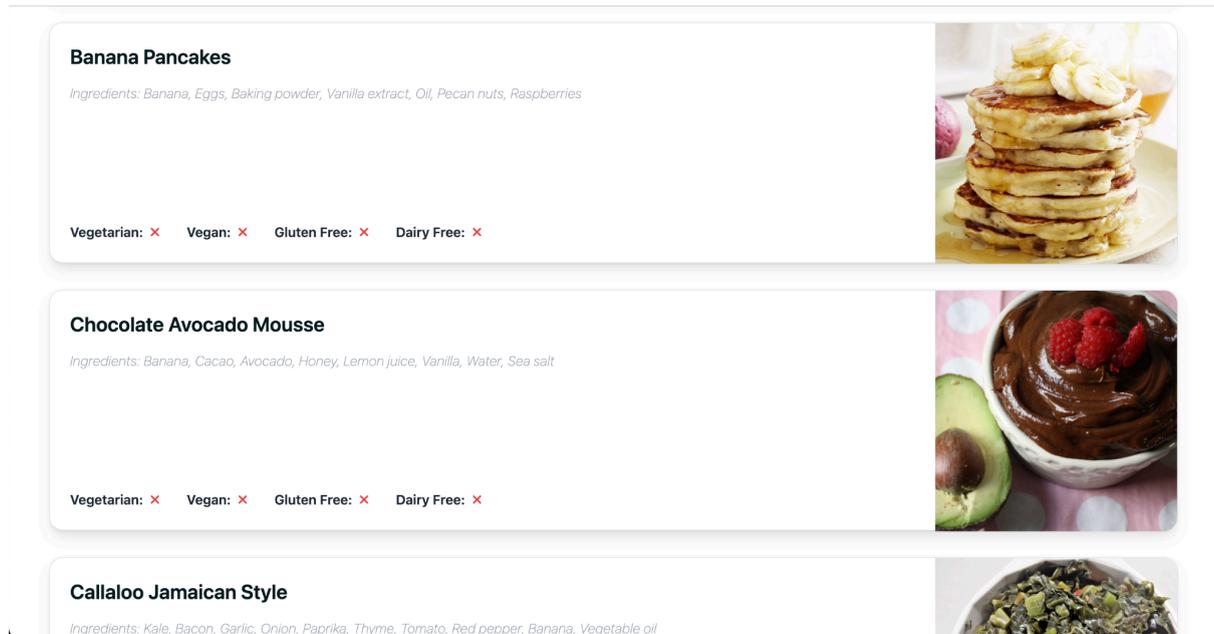


Figura 3: Listado de recetas para el ingrediente “Banana” sin ninguna restricción alimenticia.

En esta vista, el usuario puede clickear en alguna receta y así ser dirigido al detalle de la misma, donde puede ver los ingredientes y las instrucciones. En la Figura 4 se muestra en detalle la receta de “Banana Pancakes” que fue clickeada por el usuario.

 Back

Banana Pancakes



Ingredients

- Banana
- Eggs
- Baking powder
- Vanilla extract
- Oil
- Pecan nuts
- Raspberries

Vegetarian: Vegan: Gluten Free: Dairy Free:

Instructions

1. In a bowl, mash the banana with a fork until it resembles a thick purée
2. Stir in the eggs, baking powder and vanilla
3. Heat a large non-stick frying pan or pancake pan over a medium heat and brush with half the oil
4. Using half the batter, spoon two pancakes into the pan, cook for 1-2 mins each side, then tip onto a plate
5. Repeat the process with the remaining oil and batter
6. Top the pancakes with the pecans and raspberries

Figura 4: Se muestra en detalle la receta de “Banana Pancakes”.

Trabajo futuro

Se destacan varios puntos a mejorar:

- Integrar nuevas API 's, en este caso tendría sentido que estas sean de pago, lo que permitirá obtener un conjunto mayor de datos, tanto en volumen como en calidad. Se espera poder contar con mayor información para presentar, como videos, reseñas, detalles nutricionales, tiempo de cocción.
- Optimizaciones con Elasticsearch, se realizó con total éxito la integración al proyecto, pero son muchas las funcionalidades adicionales que se pueden adoptar, en un futuro se esperaría poder trabajar con estas, como por ejemplo Fuzzy Search. También resulta interesante el poder realizar búsquedas más complejas, por frases, por más de un ingrediente, entre otros.
- Mejora en la ingestión de datos, utilizar tecnologías más aptas para esto tales como Pandas.
- Implementación de usuarios y manejo de varios idiomas, esto podría requerir de una traducción completa de las recetas en caso de no contar con alguna API que proporciona los resultados en el idioma necesitado.

Conclusiones

Se considera luego de finalizado el proyecto tener un resultado satisfactorio, se cumplió con la consigna planteada, aprendiendo en la práctica conceptos de recuperación e integración de información y su aplicación en un problema real. Se utilizaron nuevas herramientas de la mano con estos conceptos que facilitan varios procesos, se destacan Docker, que facilitó el desarrollo en gran medida al estar trabajando en diferentes entornos, y Elasticsearch.

El proyecto a su vez permitió trabajar para la mayoría del grupo con un nuevo lenguaje de programación.

Y por sobre todo, se aprendió a cuando decir que un proyecto no es rentable o factible y animarse a hablarlo para ver como se continúa, en este caso cambiando de proyecto con la aprobación de la profesora.

Bibliografía

1. Documentación de Docker
<https://docs.docker.com>
2. Documentación de Ruby on Rails
<https://guides.rubyonrails.org>
3. Documentación de Elasticsearch
<https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>
4. Documentación de Spoonacular
<https://spoonacular.com/food-api/docs>
5. Documentación de TheMealDB
<https://www.themealdb.com/api.php>