

restoApp

Informe final

Recuperación de información y recomendaciones en la web 2017



Restoapp

Grupo 11

Nicolás Zeballos - CI: 4.966.740-7 - mail: nicozeba@gmail.com

Belén Saá - CI: 5.214.378-5 - mail: belensaacastro@gmail.com

Índice

Índice	1
Introducción	2
Problema	2
Enfoque de la solución	2
Diseño del sistema	3
Implementación	5
Solución desarrollada	5
Desafíos	5
Evaluación	6
Trabajo a futuro	7
Referencias	7

1. Introducción

En los últimos años, el uso de las tecnologías web ha sido aplicado en distintos rubros para poder brindar información a un público masivo. Un ejemplo de este fenómeno son los locales de comida.

“PedidosYa” ha sido un sistema pionero en el área de los servicios de comida, brindando la posibilidad de permitir al usuario solicitar un delivery sin necesidad de tener el número telefónico del local. Pero lo que se ha notado es la falta de aplicaciones y páginas web que brinden información global sobre distintos locales de comida a los que se pueda ir a comer.

Por lo tanto, el objetivo del grupo es poder brindar de forma amigable y sencilla al usuario una manera de consultar y visualizar información sobre locales de comida a través de una misma página web.

2. Problema

Se desea desarrollar un sistema que permita buscar de forma amigable al usuario locales de comida aplicando distintos filtros para facilitar la búsqueda.

La motivación surge debido a que se encontraron páginas en las cuales hay mucha información, pero la forma de búsqueda resulta engorrosa para cualquier usuario, ya que no es clara la navegación, lo que puede producir que el usuario se frustre fácilmente y decida cambiar de sitio.

3. Enfoque de la solución

Se tomaron las siguientes consideraciones al momento de desarrollar el sistema:

- Los “locales de comida” se definen como cualquier local que brinde servicio gastronómico de atención al público en el mismo local.
- Los locales de comida utilizados están ubicados únicamente en la ciudad de Montevideo.
- Las páginas utilizadas como fuente de información son:
 - www.saliracomer.com
 - www.restaurantmontevideo.com
- Se tuvo que buscar una manera de facilitar la búsqueda de información almacenada a partir de filtros que sean amigables con los usuarios.

La información que muestra el sistema es:

- Nombre del local de comida.
- Barrio de los locales de comida.
- Tipos de comida que se sirven en el local de comida.
- Rangos de precios en el local de comida.
- Servicios del local de comida (pagos con tarjeta de crédito, parking, wi-fi, etc.).
- Horario de atención.

El usuario para la búsqueda podrá aplicar la cantidad de filtros deseados en las siguientes categorías:

- Barrio.
- Tipos de comida.
- Rangos de precios.
- Servicios.

4. Diseño del sistema

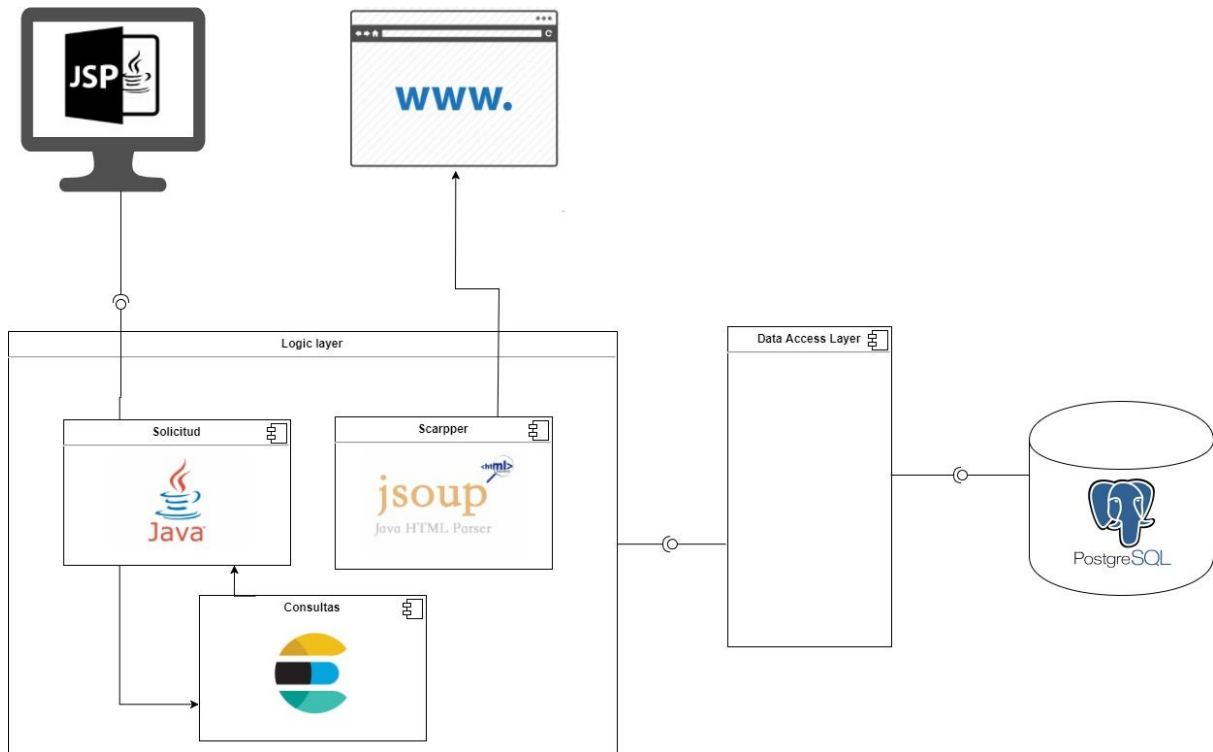
Las tecnologías utilizadas fueron las siguientes:

- PostgreSQL como motor de base de datos.
- Java como lenguaje backend.
- JSoup como scraper.
- JSP y Servlets, principalmente para el desarrollo de frontend.
- Elasticsearch para realizar búsquedas.

Las componentes del sistema son:

- Interfaz web front end.
 - Utilizada para mostrar los datos que el usuario consulta.
- **Capa “Logic Layer”**
 - Cuenta con la lógica del sistema, se divide las siguientes componentes:
 - **Scraper:** Componente utilizada para tomar los datos de las páginas web e ingresarlos en la base de datos.
 - **Consultas:** Componente para ingresar los datos a elasticsearch comunicándose con la base de datos, y resolviendo las peticiones de la componente **Solicitud**.
 - **Solicitud:** Componente utilizada para recibir las consultas del cliente, procesarlas para enviar a la componente **Consultas** la petición y devolver al frontend el resultado obtenido, mediante un nuevo procesamiento.
- **“Data Access Layer”**
 - Componente utilizada para comunicar la capa **“Logic Layer”** con la **base de datos**.
- **Motor de base de datos.**

A continuación se muestra un diagrama en que se conectan las tecnologías con las componentes:



5. Implementación

5.1. Solución desarrollada

Los pasos que se siguieron para realizar la implementación de la solución fueron los siguientes:

- Se investigaron los código fuente de ambas páginas, para saber cómo poder obtener los datos deseados extrayendo los mismo a través de las etiquetas HTML.
- Utilizando Jsoup como herramienta de web scraping, la recolección de la información se realizó una única vez y luego se trabajó con los datos almacenados en la base de datos.
- Al tener los datos recolectados a partir del scrapper en la base de datos, se genera una consulta a la misma para traerlos, y dividiendo los distintos valores de cada columna, se ingresan los datos bajo el índice “restoapp” y el tipo “restos” a elasticsearch.
- Ahora al momento de recibir una consulta desde frontend, se toma cada filtro ingresado y se genera la consulta que será enviada a elasticsearch. El mismo retornará el resultado en forma de “JSON”.
- El “JSON” obtenido como resultado se procesa y se retornan los datos a la web.

5.2. Desafíos

Los desafíos a los que se tuvo que enfrentar el equipo para poder implementar la solución al problema fueron:

- La recolección y el filtrado de la información de las páginas que se utilizarán fue compleja debido a que tienen distinto formato para mostrar la información:
 - Para solucionar este inconveniente se utilizó la librería de Java **Jsoup**, la cual es un parser Java para HTML que proporciona una API muy conveniente para la extracción y manipulación de datos. A su vez, es una herramienta gratuita y open source por lo que hay una basta información para su utilización en la web.
 - De todas formas, si bien la herramienta facilita la extracción, el modo en que está implementada una de las webs a scrapear, dificultó la tarea, no permitiendo recolectar la totalidad de la información.
- Obtener la información de manera eficiente, debido a que se presentaba una gran dificultad al momento de realizar las consultas a la base de datos para la recuperación de la misma.
 - Para solucionar esto, se decidió utilizar la herramienta **ElasticSearch**, que facilita las búsquedas por medio de creación de índices, en los cuales se ingresan “documentos” de cierto tipo, en nuestro caso los

locales de comida. Su utilización permite buscar prefijos, palabras exactas, múltiples coincidencias entre campos, y diferentes mecanismos de filtrado, pudiendo seleccionar datos que explícitamente la consulta no deba contener.

- Dificultad para el desarrollo del frontend debido a que los integrantes del grupo no conocen muchas tecnologías y no se contaba con suficiente tiempo para investigación.
 - Para abordar este problema se decidió la utilizar de **JSP y Servlets**, ya que fueron las utilizados en la materia “Taller de programación” y los integrantes contaban con experiencia sobre las mismas. Estas herramientas permiten combinar código Java con código HTML, CSS y Javascript, facilitando la forma de ingresar información en la presentación web.

6. Evaluación

En cuanto a la precisión y el recall obtenidos, los locales recolectados fueron relevantes, pero no se pudo extraer la totalidad de la información de una de las páginas, por lo que la precisión de la información, si bien fue buena, no resultó como se esperaba.

A su vez, gran cantidad de restaurantes no contaban con información acerca de los servicios brindados, por lo cual, dentro de la información relevante recolectada, la efectivamente relevante, fue muy poca. Por lo tanto, el recall tampoco resultó ser el esperado.

Si bien se pudieron recopilar los datos de varios locales de comida, se puede observar la falta de información a nivel de la web, dado que la mayoría de los restaurantes no contaban con información suficiente que permita nutrir y acompañar a la idea principal del proyecto. De todas formas, se logró extraer la información de manera automatizada y se adquirieron conocimientos en las herramientas y metodología de web scraping y eficiencia de búsquedas.

A pesar de estas consideraciones referidas a la precisión y al recall, se puede afirmar que el grupo tuvo un resultado satisfactorio, dado que este producto se toma como un prototipo base para poder continuar indagando las diferentes temáticas de la recuperación de datos en la web.

Con esta web implementada, se espera poder alcanzar esta y más información a la gente para que pueda degustar de distintos platos por toda la ciudad de Montevideo.

7. Trabajo a futuro

Para obtener mayor precisión en la información de cada local, se pretende investigar las páginas de cada local. Por lo tanto, queda pendiente la investigación en la web de páginas de varios locales de comida, así como de diferentes API's que brinden información útil (como por ejemplo "Facebook"). Complementando a esto, se pretende agregar las recomendaciones, noticias y promociones de cada local en tiempo real, para hacer llegar de forma dinámica la información a todos los usuarios de la aplicación.

A su vez, se pretende desarrollar una aplicación móvil la cual brinde la información sobre los locales de comida. Esta fue la idea desde un principio, pero se tomo la iniciativa de atacar el problema brindando una interfaz web para poder dedicar más tiempo a la investigación de las tecnologías relacionadas al procesamiento de los datos.

También, como complemento de la interfaz móvil, se busca implementar la geolocalización de los locales mostrandolos en un mapa. Esta característica pretende mostrar a los usuarios de forma más amigable la información referida a la ubicación de los locales, y a su vez presentar funcionalidades que permitan mostrar locales más cercanos a la ubicación actual del usuario, entre otras características que las herramientas referidas a esta temática permitan.

8. Referencias

- <https://jsoup.org/>
- <https://www.elastic.co/>
- <https://www.postgresql.org/>
- <https://www.java.com/es/>
- <https://www.tutorialspoint.com/jsp/>