

RECUPERACIÓN DE INFORMACIÓN Y RECOMENDACIONES EN LA WEB

CURSO 2021

GRUPO 3

Informe final

Autores

Juan Andrés Baz

juan.baz.pera@fing.edu.uy

Bruno Bradach

bruno.bradach@gmail.com

Martín Feldman

martin.feldman@fing.edu.uy

Fernando Rabago

fernando.rabago@fing.edu.uy

Introducción	3
Motivación y descripción del problema	3
Herramientas utilizadas	4
NodeJS y ReactJS	4
Elasticsearch	4
Puppeteer	4
Enfoque de la solución	4
Diseño e implementación	5
Arquitectura	5
Scraping	5
Motor de búsqueda	6
API	6
Funcionalidades y uso	7
Evaluación y resultados	11
Conclusiones	11
Trabajo Futuro	12
Referencias	13

Introducción

Un problema común que surge al organizar un evento, fiesta o reunión es ver cuánto, cómo y dónde comprar la bebida. Una vez que se calcula la cantidad necesaria de cada bebida todavía quedan por decidir la marca, presentación, el tipo específico de dicha bebida y dónde comprarla. Actualmente en el mercado se pueden encontrar muchas opciones disponibles con diversos precios, los cuales también varían dependiendo de la tienda donde se realice la compra. Generalmente frente a estas situaciones se quiere encontrar la alternativa más económica entre las posibles para así optimizar el dinero gastado en este apartado.

Dado que cada tienda tiene su propio sitio web con su catálogo de precios, actualmente es necesario buscar en cada sitio por separado y hacer las comparaciones uno mismo.

Para este trabajo decidimos enfocarnos en la resolución de este problema, utilizando herramientas adquiridas durante el curso, para crear una aplicación web que permita obtener información de diferentes sitios web donde se proveen bebidas, con su respectivo precio y descripción. Dicha información va a ser presentada al usuario de forma amigable y centralizada en un sitio web de forma que este pueda comparar y conseguir su bebida favorita al menor precio posible, evitando tener que ingresar a cada sitio para registrar los precios. El sitio web desarrollado tendrá diversas categorías, un buscador y filtros.

En este informe presentaremos el problema a resolver, describiremos la aplicación diseñada y las tecnologías utilizadas para su implementación, dejaremos ejemplos de su funcionalidad y su uso, evaluaremos su funcionamiento y mencionaremos posibles mejoras y trabajos a futuros a realizar.

Motivación y descripción del problema

La motivación de este proyecto surge por una situación en la cual todos los integrantes del equipo se han enfrentado en repetidas ocasiones, al tener que conseguir ciertas bebidas a un precio accesible para todos los participantes de un reunión o fiesta; por lo tanto se nos ocurrió poder tener acceso rápido e inmediata a esa información de precios y bebidas en diferentes lugares y poder tomar la mejor decisión posible en esa situación.

Herramientas utilizadas

NodeJS y ReactJS

La elección de estas herramientas viene dada por sus buenas capacidades a la hora de crear aplicaciones web basadas en APIs Rest, además de ser ya conocidas y manejadas por los integrantes del equipo.

Elasticsearch

Se eligió Elasticsearch ya que entendimos que era altamente recomendable para el almacenamiento y procesamiento de grandes volúmenes de datos. Además, ofrece un lenguaje de consulta muy extenso y posibilidad de realizar búsquedas por palabras clave dentro de los documentos a la hora de efectuar las consultas, lo cual era muy importante para nosotros.

Puppeteer

Es una biblioteca de Node que proporciona una API de alto nivel para controlar el navegador Chromium o Chrome a través del protocolo DevTools. Además se utilizó esta librería ya que es conocida por algunos integrantes del equipo y es bastante fácil y rápida de usar.

Enfoque de la solución

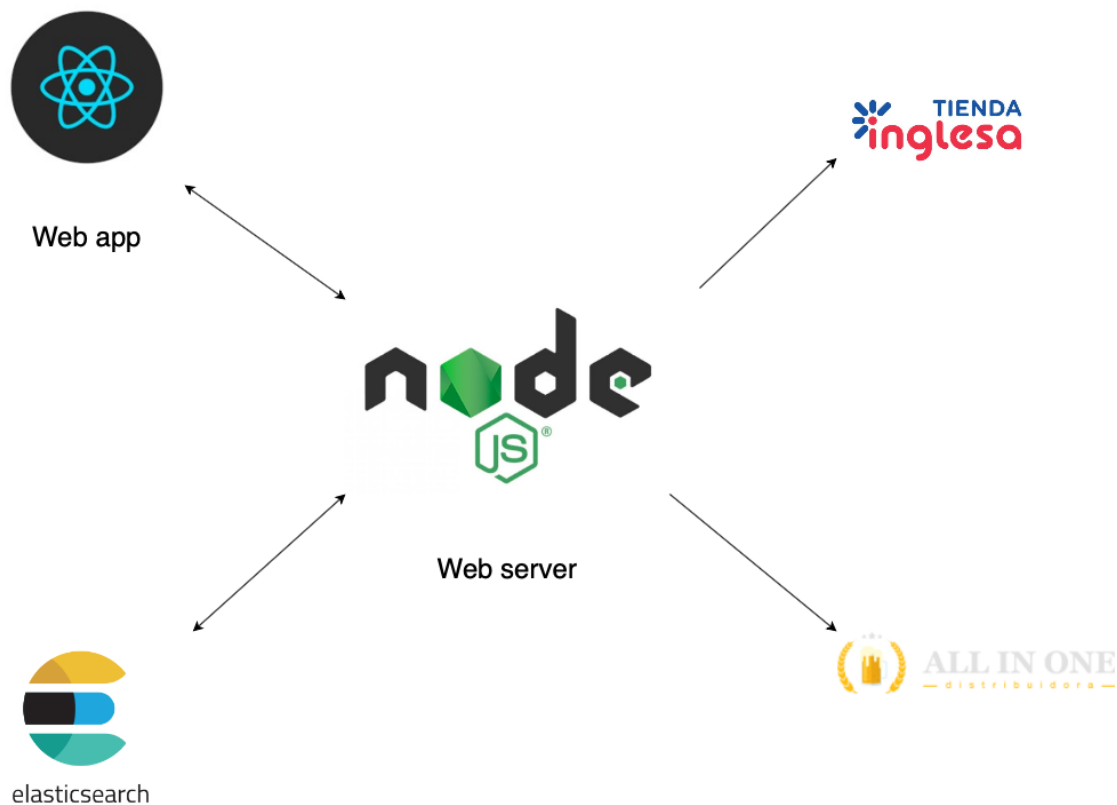
El alcance del proyecto fue pactado entre el equipo de desarrollo y la docente de monitoreo en abarcar cerveza y whisky (en cuanto a bebidas) y las tiendas Tienda Inglesa y Distribuidora Bebidas

El enfoque de la solución es desarrollar un sistema de búsqueda y recuperación de información, el cual obtendrá los datos de las bebidas mencionadas desde dos tiendas seleccionadas y los presenta al usuario de una forma intuitiva en un sitio web. Los datos a mostrar son: precio, imagen, nombre del producto y tienda donde se vende, además del link a la página del producto en el sitio web de la tienda. Se incluye una barra de navegación para buscar las bebidas por su nombre. A su vez, se agregaron filtros que brindan más funcionalidades a la aplicación haciéndola mucho más útil, además de mejorar la experiencia del usuario al utilizarla.

Diseño e implementación

Arquitectura

La siguiente imagen muestra la arquitectura del sistema implementado:



Los principales componentes que conforman el sistema son:

- Módulo encargado de recuperar los datos de las tiendas para posteriormente almacenarlos en el servidor de ElasticSearch (se usó Node.js).
- Aplicación web que consume los datos almacenados en ElasticSearch y permite la aplicación de filtros y búsquedas sobre las bebidas. Esta fue desarrollada con React JS.

Scraping

Esta técnica fue la utilizada para extraer la información de las tiendas. La misma se basa en simular la navegación de un humano en la web y acceder al contenido HTML mediante HTTP o, como en este caso particular, mediante un navegador headless utilizando la librería Puppeteer.

Se implementó un algoritmo que toma una lista de URLs de un archivo de configuración y para cada una obtiene la información del listado de productos presentes en dicha URL, además de navegar las páginas siguientes en caso de contar con paginación.

A continuación se presenta un ejemplo de dicho archivo de configuración.

```
{
  "name": "tienda inglesa",
  "pages": [
    {
      "url": "https://www.tiendainglesa.com.uy/busqueda?Patagonia",
      "category": "beer",
      "store": "tienda inglesa",
      "brand": "patagonia"
    },
    ...
  ]
}
```

El algoritmo de scraping se basa en selectores CSS para obtener la información del HTML, los cuales son previamente obtenidos manualmente analizando el sitio web de la tienda. El diseño del módulo de scraping permite la incorporación de otras tiendas para que formen parte de la comparación, agregando nuevos conjuntos de selectores para cada tienda.

Una vez obtenida la información, se almacena de forma auxiliar en una estructura de datos que será utilizada para indexar la información en Elasticsearch.

Motor de búsqueda

En Elasticsearch se creó un índice *products* que indexará todos los productos. Algunas de las propiedades de dicho índice son *category*, *store* y *brand*, que son utilizadas para implementar los filtrados en la aplicación web. De esta forma, los filtros en la aplicación web son generados automáticamente a partir de la información de los productos indexados.

Tanto el indexado como la búsqueda de productos son realizados desde la API utilizando el cliente de NodeJS provisto por Elasticsearch.

API

La API provee dos operaciones posibles:

- *GET /products*: utilizada para obtener el listado de productos en base a los parámetros de búsqueda y filtrado especificados en la request.
- *GET /filters*: utilizada para obtener el listado de filtros disponibles.

Funcionalidades y uso

Las principales funcionalidades provistas por la aplicación son la barra de búsqueda, permitiendo al usuario localizar un ítem específico de su interés, y el conjunto de filtros. En una misma consulta se pueden aplicar varios filtros en conjunto, permitiendo así realizar consultas más complejas/específicas a los usuarios que así lo dispongan.

Se presentan a continuación algunas imágenes de la aplicación web funcionando.

Al ingresar a la página, se muestra una pantalla inicial la cual contienen un conjunto de bebidas sin filtrar:

WEBIRRA

Bebidas









Filtros

Tiendas

- Tienda Inglesa
- Distribuidora Bebidas

Marcas

- Johnnie Walker
- Patagonia
- Volcanica
- Patricia
- Zillertal
- Chivas Regal
- Dab
- Heineken
- Jack Daniels
- Estrella Galicia

 <p>\$</p> <p>Cerveza Estrella Galicia Bot 355cc</p> <p>Distribuidora Bebidas</p>	 <p>\$ 59</p> <p>Cerveza Estrella Galicia 0,0 Alcohol bot 250cc</p> <p>Distribuidora Bebidas</p>	 <p>\$</p> <p>Cerveza Estrella Galicia Lata 269cc por 12 Unidades</p> <p>Distribuidora Bebidas</p>	 <p>\$ 84</p> <p>Cerveza Estrella Galicia Sin Gluten bot 330cc</p> <p>Distribuidora Bebidas</p>
 <p>\$ 48</p> <p>Cerveza Estrella Galicia Lata 350ml</p> <p>Distribuidora Bebidas</p>	 <p>\$ 54</p> <p>Cerveza Heineken Botella 330 ml</p> <p>Distribuidora Bebidas</p>	 <p>\$ 72</p> <p>Cerveza Heineken Lata 473 cc (Cantidad mínima 12 uds)</p> <p>Distribuidora Bebidas</p>	 <p>\$ 103</p> <p>Cerveza Heineken Lata 710 ml (Cantidad mínima 4 uds)</p> <p>Distribuidora Bebidas</p>

En la siguiente imagen se puede ver el resultado de la aplicación de algunos filtros sobre el conjunto de bebidas perteneciente al sistema:

WEBIRRA

Bebidas











Filtros

Tiendas

- Tienda Inglesa
- Distribuidora Bebidas

Marcas

- Johnnie Walker
- Patagonia
- Volcanica
- Patricia
- Zillertal
- Chivas Regal
- Dab
- Heineken
- Jack Daniels
- Estrella Galicia

 <p>\$ 37</p> <p>Botella Cerveza Patricia 340 cc Retornable</p> <p>Distribuidora Bebidas</p>	 <p>\$ 61</p> <p>Lata Cerveza Patricia 473 cc Oferta</p> <p>Distribuidora Bebidas</p>	 <p>\$ 113</p> <p>Botella Cerveza Patricia Litro Retornable</p> <p>Distribuidora Bebidas</p>	 <p>\$ 147</p> <p>Cerveza PATRICIA Botella Retornable 960 ml</p> <p>Tienda Inglesa</p>
 <p>\$ 84</p> <p>Cerveza PATRICIA Lata 473 ml</p> <p>Tienda Inglesa</p>	 <p>\$ 85</p> <p>Cerveza PATRICIA Dunkel Lata 473 ml</p> <p>Tienda Inglesa</p>	 <p>\$ 85</p> <p>Cerveza PATRICIA Porter lata 473 ml</p> <p>Tienda Inglesa</p>	 <p>\$ 147</p> <p>Cerveza PATRICIA Porter Botella Retornable 960 cc</p> <p>Tienda Inglesa</p>
			

Si se quisiera buscar un ítem en específico se puede utilizar el buscador ubicado en la parte superior de la pantalla:



Bebidas













Filtros

Tiendas

- Tienda Inglesa
- Distribuidora Bebidas

Marcas

- Johnnie Walker
- Patagonia
- Volcanica
- Patricia
- Zillertal
- Chivas Regal
- Dab
- Heineken
- Jack Daniels
- Estrella Galicia

 <p>\$ 141 Cerveza Zillertal IPA 1L Distribuidora Bebidas</p>	 <p>\$ 71 Cerveza Zillertal Ipa Lata 473cc Distribuidora Bebidas</p>	 <p>\$ 163 Cerveza BIZARRA IPA Suave 500ml Tienda Inglesa</p>	 <p>\$ 145 Cerveza VOLCANICA Belgian Ipa 500 ml Tienda Inglesa</p>
 <p>\$ 145 Cerveza VOLCANICA American IPA 500 ml Tienda Inglesa</p>	 <p>\$ 159 Cerveza American Ipa VOLCANICA 500 ml Tienda Inglesa</p>	 <p>\$ 94 Cerveza ZILLERTAL IPA Lata 473 ml Tienda Inglesa</p>	 <p>\$ 178 Cerveza ZILLERTAL IPA Botella 1 L Tienda Inglesa</p>
			

Evaluación y resultados

Se logró construir una aplicación web con cortos tiempos de respuesta y agradable interfaz gráfica, la cual creemos que es intuitiva y fácil de utilizar para cualquier usuario básico de la web. Realiza correctamente su trabajo, recopilando de los tiendas descritas las bebidas seleccionadas y mostrandose las al usuario en un formato minimalista.

La evaluación de la solución fue satisfactoria, ya que la web brindó buenos resultados a las búsquedas realizadas. Dichas pruebas cubren necesidades comunes como buscar una marca en particular, solamente latas, solamente botellas de litro, etc.

Conclusiones

Creemos que el proyecto realizado fue enriquecedor para el equipo, ayudando a comprender las posibilidades de recuperación y recopilación de información en la web para su almacenamiento y consecuente procesamiento.

El desafío de aprender una nueva tecnología, como fue el caso de *Elasticsearch*, siempre supone un desafío que conlleva riesgos. Sin embargo, el resultado final fue satisfactorio, pues se consiguieron superar los desafíos encontrados para terminar obteniendo el producto deseado. Del lado de React y Node son tecnologías que la mayoría de los integrantes del grupo utilizaron en su momento y no significó un costo de tiempo en el aprendizaje si no más que nada una facilidad en ese sentido.

No tenemos duda de que el sitio web desarrollado sería de gran interés para una gran cantidad de usuarios, principalmente aquellos comprendidos entre los 18 y 40 años, quienes se enfrentan a esta elección de donde comprar bebidas para sus eventos cotidianamente. A su vez este público está habituado a interactuar con sitios web y suele buscar el mejor precio entre las diferentes alternativas.

Habiendo finalizado el proyecto el equipo considera que se cumplió tanto con las expectativas como con el alcance del mismo, cumpliendo con el problema planteado en la introducción de este documento. Esto se justifica con haber realizado todo lo planificado en la presentación inicial. Por otro lado, la opinión del equipo es que el diseño de la página es amigable y fácil de usar para todo tipo de usuario final acostumbrado a navegar en la web.

A pesar de esto, pensamos que antes de una primera liberación de la aplicación se debería trabajar más en la misma de forma que brinde una experiencia más completa al usuario.

Notamos varios aspectos en los cuales se puede mejorar/enriquecer la aplicación, que serán detallados en la siguiente sección.

Trabajo Futuro

Creemos que esta aplicación tiene un buen potencial para expandir tanto los datos como las funcionalidades del sistema de forma de generar una plataforma más completa en estos sentidos.

- El sistema desarrollado carga los datos una única vez, creemos que el scrapping debería ser realizado de forma automatizada para así tener información actualizada en Elasticsearch. Como los precios de estas tiendas no son actualizados constantemente el scrapping podría ser realizado cada 4 o 6 horas de forma de no consumir recursos de manera innecesaria.
- El sistema actualmente solo tiene datos de cerveza y whisky de las tiendas Tienda Inglesa y Distribuidora Bebidas. Creemos que para que el mismo sea realmente útil se deberían agregar más tiendas y bebidas de forma de ampliar el volumen de datos presentado al usuario. Esta expansión estará condicionada por la estructura de los sitios web de las tiendas, siendo necesaria una estructura HTML que permita al módulo de scrapping identificar la información correctamente.
- Añadir una sección donde se ingrese una lista de productos deseados y que la web recomiende la mejor compra posible.
- Se podrían implementar una gran variedad de filtros. Entre los que destacamos están:
 - Rango de precios.
 - Tiendas con envío gratis.
 - Tipo de bebida (Fermentadas, Destiladas, Licores, etc)
 - Tipo de cerveza (Ale, Lager, Porter, etc)
 - Apto para celíacos, sin alcohol.
 - Tipo de fabricación (Industrial, Artesanal)
 - Origen (Nacional, Extranjera)
 - Presentación de la bebida (Lata 500 ml, botella 1 L, etc)
- Herramienta que permita calcular la relación cantidad/costo de las bebidas.
- Sección con las ubicaciones y horarios de las tiendas disponibles.
- Añadir sección de promociones donde se advierta si hay descuentos en las tiendas de la plataforma. Dependerá de si los sitios web proveen esta información de una forma interpretable mediante scrapping.
- Se podría generar una Progressive Web App para utilizar en dispositivos móviles.
- Implementar un carrito de compra para agregar los productos, el mismo estaría particionado según las tiendas de las cuales se escogieron las bebidas.
- Mejorar la búsqueda para soportar búsquedas de substrings más complejas.
- Información sobre si el envase está incluido en el precio.

Referencias

Elasticsearch web: <https://www.elastic.co/>

NodeJS web: <https://nodejs.org/en/>

ReactJS web: <https://reactjs.org/>

Web Scraping: https://es.wikipedia.org/wiki/Web_scraping

Puppeteer: <https://github.com/puppeteer/puppeteer>