

UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

RECUPERACIÓN DE INFORMACIÓN Y RECOMENDACIONES EN LA WEB



Informe final

Grupo 16

Rodrigo Negreyra CI: 5.098.239-5
Santiago Fernandez CI: 5.046.100-0
Juan Tassino CI: 5.092.368-6

Docente
Libertad Tansini

Índice

Índice	2
Problema	3
Enfoque de la solución	3
Opción 1:	3
Opción 2:	4
Diseño	5
Tecnologías	5
Implementación	6
Problemas encontrados	7
Trabajo Futuro	8
Conclusiones	9
Referencias	10

Problema

Mediante la búsqueda de posibles ideas que podrían ser útiles para explorar los temas del curso, nos dimos cuenta al charlar entre los integrantes del grupo, que varios de nosotros estábamos en búsqueda de un apartamento para mudarnos.

Existen varias opciones de búsqueda de apartamentos en internet que brindan muchas opciones parecidas, pero al mismo tiempo existen muchas opciones que no se encuentran en las otras páginas.

Esto termina generando que la búsqueda de apartamentos sea bastante engorrosa, teniendo que buscar en distintas páginas, cada una con su propia interfaz y su propia forma de hacer las búsquedas.

Enfoque de la solución

Se propone la implementación de una web que centralice todas estas búsquedas y ofrezca, en un mismo lugar, el resultado de realizar una búsqueda en varias aplicaciones, así teniendo un paneo general de la disponibilidad que existe en el mercado.

Para la prueba de concepto se decidió tomar información de las siguientes páginas web:

- Info Casas ¹
- Mercado Libre ²

Opción 1:

Se realiza la búsqueda de la información “on the fly”, replicando dicha búsqueda en todas las webs consumidas, obteniendo las respuestas y transformándose de forma dinámica para normalizar la información. Esto requiere de herramientas como el scraping web y la posibilidad de ejecutar consultas a API's de las webs de forma eficiente, para no tener un costo de tiempo alto al usuario.

Ventajas:

- Información actualizada con los últimos datos publicados.
- Menor costo de infraestructura, al no requerir de una base de datos.

Desventajas:

- Posibilidad de tener tiempos de respuesta muy altos para ser viable una experiencia de usuario placentera.
- Limitaciones a la hora de realizar las consultas por los metadatos que acepten ambos proveedores en sus consultas (ej: filtros o búsquedas por texto deberían ser posibles en ambos proveedores).

¹ <https://www.infocasas.com.uy/>

² <https://listado.mercadolibre.com.uy/montevideo>

Opción 2:

Se realiza un preprocesamiento de un dataset obtenido de ambos proveedores, generando una indexación como la dada en el curso, y guardando los datos en una base de datos con búsquedas indexadas (ej: elasticsearch) que permita hacer búsqueda de los datos de forma rápida y eficiente.

Ventajas:

- Posibilidad de tener control sobre la base de datos y hacer búsquedas relativamente eficientes en el tiempo.
- Posibilidad de indexar datos diferentes a los metadatos permitidos por los proveedores a la hora de consultar dinámicamente. Ej: se podrán realizar búsquedas por texto más complejas.

Desventajas:

- Información desactualizada con los últimos datos presentes en las web de los proveedores (pueden mostrarse publicaciones ya cerradas, o no mostrarse nuevas, que sean más recientes al último dataset de datos preprocesado)
- Existe el costo y complejidad de mantener, así como actualizar la base de datos requerida para esta opción.

Solución

Luego de analizar las dos opciones disponibles, realizamos tests de verificación y encontramos que el tiempo de respuesta al consultar las webs propuestas de forma dinámica tienen tiempos aceptables, lo cual eliminaría esa desventaja de la opción 1.

También nos parece que realizar búsquedas complejas no es necesario debido a que el caso de búsqueda de inmuebles generalmente tiene parámetros bastante estáticos de búsqueda, como pueden ser: nro habitaciones, nro de baños, etc.

Finalmente, creemos que el hecho de tener la información actualizada con los últimos datos disponibles es un requerimiento necesario para nuestro caso de uso, ya que la publicación de inmuebles es muy dinámica, día a día se dan de baja publicaciones por ya haberse alquilado, y aparecen muchos inmuebles nuevos.

Por todo lo mencionado, decidimos que la opción más indicada para nuestro caso es la opción 1.

Diseño

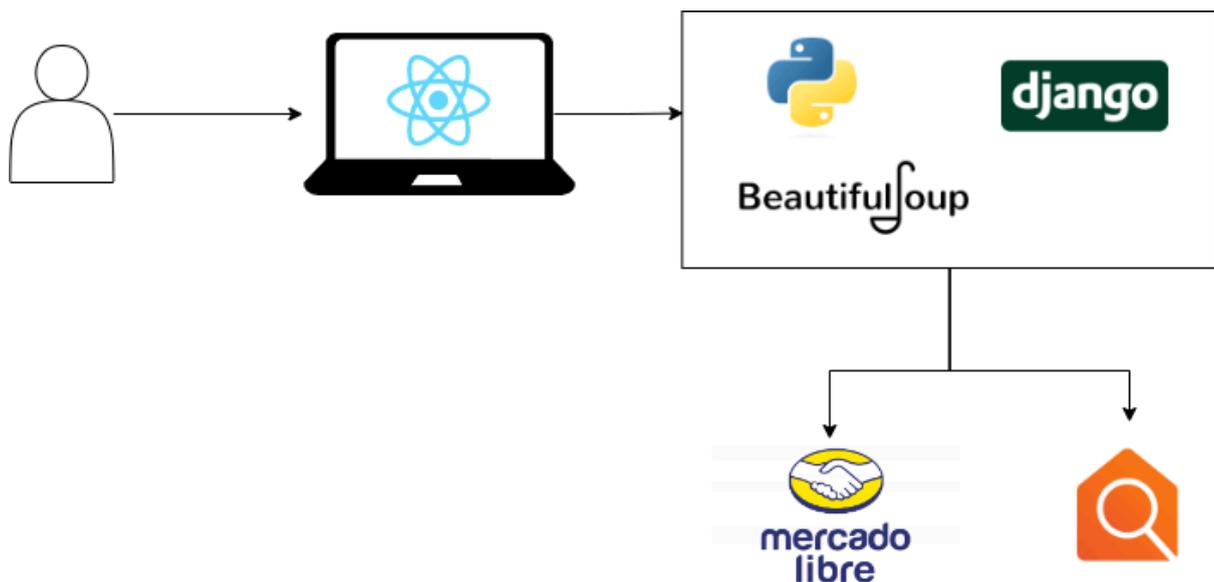
Para nuestra prueba de concepto, decidimos tener un frontend web que permita seleccionar los diferentes filtros, y un backend que haga las consultas a las plataformas de publicación de inmuebles.

El frontend hará las consultas al backend y este le devolverá la información normalizada de cada publicación obtenida en cada plataforma, obtenida mediante una API pública con GraphQL para la plataforma de InfoCasas, y mediante scrapping de la web de resultados para Mercado Libre.

Tecnologías

Se optó por usar el lenguaje python, debido a su gran soporte y diversidad de soluciones en forma de librerías de uso público:

- BeautifulSoup (Librería utilizada para hacer scraping de páginas web)
- Django (Librería utilizada para hacer apis)
- React (Lenguaje utilizado para frontend web)



Implementación

Al verificar ambos proveedores, visualizamos un conjunto de parámetros por los que se podría consultarse a ambos de forma dinámica, y a terminos de nuestra prueba de concepto, decidimos elegir los siguientes:

- Tipo de operación (compra o venta)
- Barrio
- Cantidad de dormitorios
- Cantidad de baños
- Rango de precio.

A modo de nuestra prueba de concepto, consideramos suficiente poder visualizar los siguientes datos de cada publicación:

- Título
- Imagen principal de la publicación
- Precio y moneda
- Breve descripción
- Enlace hacia la publicación

En el frontend, al mostrar los resultados, se permite ordenar las publicaciones mostradas por precio ascendente y descendente.

Pantalla inicial con filtros de búsqueda:



Pantalla mostrando los resultados de búsqueda:

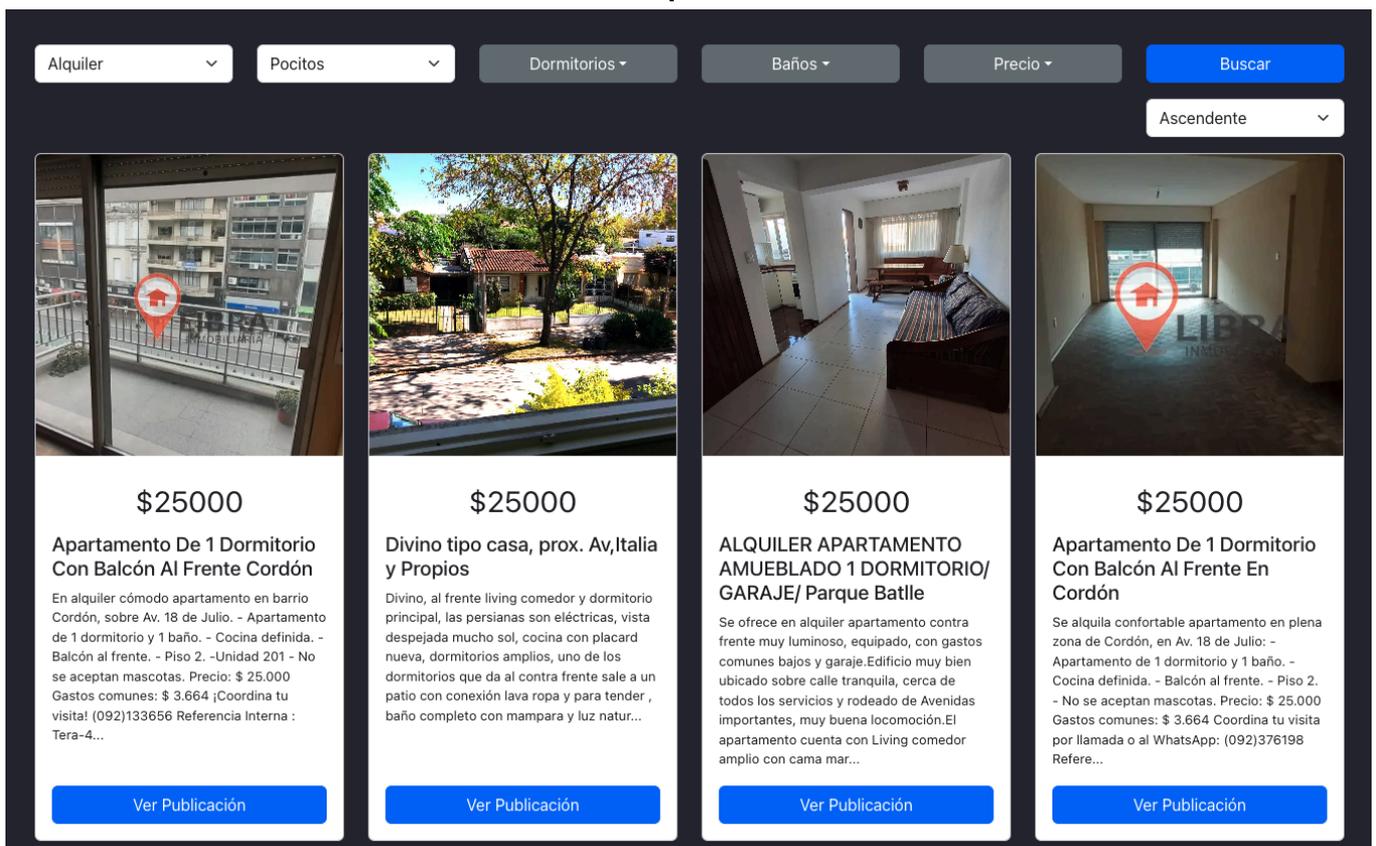


Imagen	Precio	Título	Descripción	Botón
	\$25000	Apartamento De 1 Dormitorio Con Balcón Al Frente Cordón	En alquiler cómodo apartamento en barrio Cordón, sobre Av. 18 de Julio. - Apartamento de 1 dormitorio y 1 baño. - Cocina definida. - Balcón al frente. - Piso 2. -Unidad 201 - No se aceptan mascotas. Precio: \$ 25.000 Gastos comunes: \$ 3.664 ¡Coordina tu visita! (092)133656 Referencia Interna : Tera-4...	Ver Publicación
	\$25000	Divino tipo casa, prox. Av,Italia y Propios	Divino, al frente living comedor y dormitorio principal, las persianas son eléctricas, vista despejada mucho sol, cocina con placard nueva, dormitorios amplios, uno de los dormitorios que da al contra frente sale a un patio con conexión lava ropa y para tender , baño completo con mampara y luz natur...	Ver Publicación
	\$25000	ALQUILER APARTAMENTO AMUEBLADO 1 DORMITORIO/ GARAJE/ Parque Batlle	Se ofrece en alquiler apartamento contra frente muy luminoso, equipado, con gastos comunes bajos y garaje.Edificio muy bien ubicado sobre calle tranquila, cerca de todos los servicios y rodeado de Avenidas importantes, muy buena locomoción.El apartamento cuenta con Living comedor amplio con cama mar...	Ver Publicación
	\$25000	Apartamento De 1 Dormitorio Con Balcón Al Frente En Cordón	Se alquila confortable apartamento en plena zona de Cordón, en Av. 18 de Julio: - Apartamento de 1 dormitorio y 1 baño. - Cocina definida. - Balcón al frente. - Piso 2. - No se aceptan mascotas. Precio: \$ 25.000 Gastos comunes: \$ 3.664 Coordina tu visita por llamada o al WhatsApp: (092)376198 Refere...	Ver Publicación

Problemas encontrados

- El scraping web en la página de Mercado Libre. Al ser una plataforma polifuncional e irregular, las publicaciones tienen información estándar, que es válida para la variedad de productos que se publican en dicho portal. Al no tener la misma información que en InfoCasas, tuvimos que adaptar la pantalla de resultados para mostrar los datos de las publicaciones que nos ofrecía Mercado Libre.
- Al consumirse más de un proveedor de publicaciones, pueden existir casos en los que 2 o más proveedores tengan publicado el mismo inmueble. Reconocer y diferenciar las publicaciones es un punto que en principio lo vimos como un aspecto fundamental de la aplicación, ya que era de las principales motivaciones para la realización del proyecto. Desafortunadamente, identificar que dos publicaciones correspondan al mismo inmueble es una tarea muy compleja, y creemos que escapa a las intenciones de nuestra prueba de concepto, por lo que decidimos dejar la tarea de filtrar dichos duplicados por fuera de nuestro scope.
- Encontramos que al consultar la lista de publicaciones de Mercado Libre, al parsear el html, no encontrábamos ninguna descripción de la publicación disponible. En caso de querer tener la descripción se debería hacer una consulta a cada una de las páginas web de cada una de las publicaciones de Mercado Libre, pero eso agregaría al costo de respuesta. Por lo que optamos por generar una descripción que contiene los metadatos relevantes que podemos obtener (cantidad de baños, dormitorios, metros cuadrados)

Trabajo Futuro

Publicaciones duplicadas: Como mencionamos anteriormente existe el caso de publicaciones de distintos proveedores que refieren al mismo inmueble. Además de la posibilidad de identificarlos creemos que una posible mejora para la aplicación podría ser la funcionalidad de distinguir publicaciones de distintas plataformas que hagan referencia al mismo inmueble, y permitir comparar ambas publicaciones, por ejemplo, contrastando los precios de alquiler o venta en ambas plataformas, pudiendo elegir la opción que más convenga a la hora de la búsqueda.

Icono de proveedor: Un icono mostrando de donde proviene la publicación, brindaría fácilmente al usuario esta información.

Proveedores: Aumentar el rango de búsqueda de inmuebles, esto se podría hacer agregando más proveedores de información, como la página de el Gallito.

Parámetros de búsqueda: Filtrar por más parámetros, como cantidad de metros cuadrados, si cuenta con garage u otras facilidades como aire acondicionado, piscina, etc

Ordenamiento: Posibilidad de ordenar los resultados por otro parámetros, como las publicaciones más recientes.

Información de la publicación: Agregar otras cualidades de las publicaciones, como el carrusel con todas las imágenes del inmueble presentadas por el sitio, y la ubicación proporcionada.

Mapa: Agregar un mapa con ubicación exacta de los inmuebles para facilitar la búsqueda por zona al usuario.

Cantidad resultados: La cantidad de resultados mostrados es aproximadamente 100, acotada por las 48 publicaciones máximas que muestra Mercado Libre en su pantalla de resultados. Una opción de mejora sería aumentar esta cantidad, recorriendo las diferentes páginas de resultados de Mercado Libre, y aumentando la cantidad de resultados mostrados de InfoCasas.

Conclusiones

Creemos que logramos implementar una prueba de concepto de forma exitosa, la funcionalidad lograda cumple con las expectativas que teníamos y con las que creíamos que nos brindaría valor como usuarios a la hora de continuar nuestra búsqueda de apartamento.

Al mismo tiempo creemos que existen muchas oportunidades de mejora, y aspectos que serían necesarios para poder lanzarlo como producto al mercado. Problemas como el duplicado de publicaciones, creemos que brindaría confusión y descontento, así que mínimamente nos parece requerido el poder identificarlos y mostrar solo uno para no empeorar la experiencia de usuario.

Creemos que la decisión de explorar esta idea de concepto fue muy enriquecedora, al permitarnos indagar en tecnologías como Django, GraphQL o el scraping web.

Referencias

- [1] <https://www.infocasas.com.uy/>
- [2] <https://www.mercadolibre.com.uy/c/inmuebles#menu=categories>
- [3] <https://www.djangoproject.com/>
- [4] <https://realpython.com/beautiful-soup-web-scraper-python/>
- [5] <https://graphql.org/>