

# **Recuperación de Información y Recomendaciones en la Web**

**Informe Final - 2020**

**Agregador de Alquiler y Venta de Inmuebles**

**Grupo 2**

**Emiliano Luna - 5.191.524-0**

**Andrés Noel - 4.608.273-7**

# Índice

<b>Introducción</b>	<b>3</b>
<b>Problema</b>	<b>3</b>
<b>Enfoque de la Solución</b>	<b>3</b>
<b>Diseño</b>	<b>3</b>
Backend	4
Frontend	5
<b>Implementación</b>	<b>5</b>
<b>Funcionalidades y Uso</b>	<b>6</b>
<b>Evaluación y Resultados</b>	<b>8</b>
<b>Trabajo Futuro</b>	<b>8</b>
<b>Conclusiones</b>	<b>8</b>
<b>Referencias</b>	<b>9</b>

# Introducción

En este trabajo se busca desarrollar un sistema para recuperar información de varias fuentes diferentes y con metodologías distintas, como *scrapping* y mediante el uso de una API pública. Luego de recuperada dicha información, se la lleva a un formato común permitiendo presentarla homogéneamente al usuario final.

En este informe comenzamos describiendo el contexto del problema que se abordó y el enfoque de la solución elegida. Luego detallamos su diseño e implementación. Finalmente mostramos las funcionalidades y resultados obtenidos del sistema, y discutimos posibles mejoras como trabajo futuro.

## Problema

En estos tiempos, cuando una persona está en búsqueda de un nuevo hogar en forma de alquiler o compra, ya sea casa o apartamento, se suelen visitar a sitios web como MercadoLibre, GallitoLuis, InfoCasas, Inmuebles.com.uy, o incluso sitios web de las propias inmobiliarias. Una vez allí se seleccionan filtros que se adecúen a la búsqueda que se quiere realizar, y luego se examinan las diferentes publicaciones. Finalmente, cuando se encuentra un inmueble de interés, la persona puede contactar al autor de la publicación para, por ejemplo, coordinar una visita del lugar.

Este proceso lleva mucho tiempo ya que la elección de un nuevo hogar debe realizarse de forma cuidadosa, por lo que la persona suele acceder a varios sitios para efectuar la misma búsqueda. Por otro lado, al existir tantos sitios, una persona puede perderse oportunidades de su potencial interés por no poder visitarlos todos.

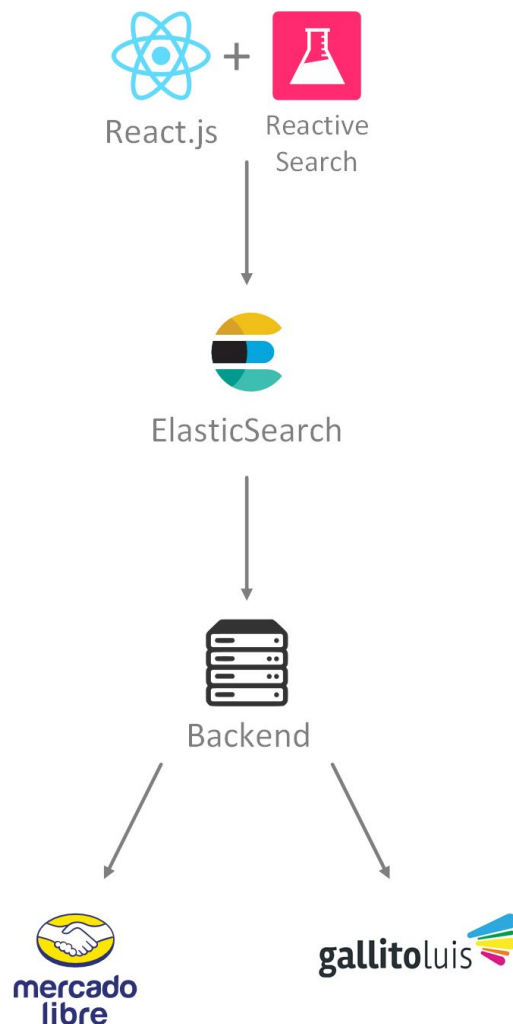
## Enfoque de la Solución

El enfoque de la solución es la de desarrollar un sistema de búsqueda y recuperación de información, que obtenga los datos de los inmuebles en alquiler o venta de MercadoLibre y GallitoLuis (los dos sitios más populares) y los presente al usuario de una forma intuitiva en un sitio web. De este modo, un usuario en búsqueda de un inmueble puede acceder a un sitio donde encontrará en una sola búsqueda la mayoría de los inmuebles en alquiler o venta del mercado.

El sistema será diseñado de forma modular para permitir la eventual incorporación de otras fuentes de datos con el fin de consolidar aún más los inmuebles disponibles en un solo lugar.

# Diseño

La siguiente imagen muestra la arquitectura del sistema implementado:



## Backend

El sistema consiste de un backend implementado en Python, el cual obtendrá los datos de inmuebles de dos fuentes, utilizando dos métodos distintos:

- > [MercadoLibre](#): se utiliza la librería oficial de MercadoLibre para Python, la cual se comunica con su API para desarrolladores.
- > [Gallito](#): se realiza un scraping de la web, particularmente en la sección de Inmuebles, utilizando BeautifulSoup

Estos datos fueron cargados en una base de datos ElasticSearch. La carga de datos se ejecuta de forma independiente al funcionamiento del sitio web y a demanda.

ElasticSearch publica una API REST la cual es accedida desde el frontend para presentar los datos al usuario.

## Frontend

El frontend fue desarrollado en React utilizando el framework ReactiveSearch el cual facilita la interacción con Elasticsearch. El usuario final puede filtrar entre diversos atributos como: Tipo de Publicación (Venta/Alquiler), Sitio Publicado (MercadoLibre/Gallito), Moneda (UYU/USD), Rango de Precio, Cantidad de Dormitorios, Rango de Superficie y Ubicación.

A su vez se pueden ordenar las publicaciones por precio o superficie de forma ascendente o descendiente.

## Implementación

### API MercadoLibre

Para utilizar la API de MercadoLibre, primero fue necesario registrar una aplicación y obtener sus credenciales.

La API utiliza el protocolo OAuth, por lo que fue necesario obtener un token de acceso y de refresco mediante una Request al endpoint de Autorización.

El token de acceso debe ser agregado a todas las requests consecuentes. Como tiene una corta expiración, debe ser actualizado mediante el token de refresco.

La API funciona de forma global para todos los sitios de MercadoLibre, por lo que es importante indicar para qué sitio (o país) se quiere realizar la solicitud.

La forma de obtener los inmuebles fue a través del endpoint de búsqueda en la categoría adecuada:

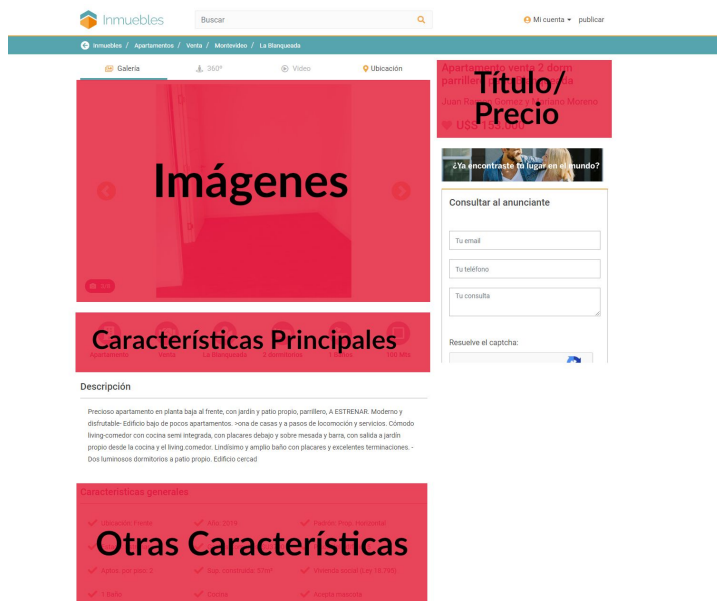
```
/sites/:sitio/search?category=:categoria&offset=xxx
```

El límite de resultados por respuesta era de 50, por lo que fue necesario realizar varias requests utilizando paginación.

### Scraping Gallito

Para obtener los datos necesarios del Gallito, se implementaron dos scripts. El primero ejecuta un loop el cual realiza una request a la página búsqueda de inmuebles, y obtiene las URLs de los resultados de la página, hasta obtener la cantidad de páginas indicadas por el usuario.

El segundo script solicita las URLs una a una y obtiene datos de cuatro secciones distintas.



Una dificultad encontrada al hacer el scraping, fue que la página carga información de forma dinámica utilizando Javascript, por lo que realizar una simple request GET devolvía un objeto HTML incompleto. Para solucionar esto se utilizó la librería de Python requests-html, la cual devuelve el objeto luego de que los scripts Javascript terminen de cargar.

Para ambos métodos, los datos son analizados y guardados en un objeto de tipo Inmueble el cual sirve para estandarizarlos. Luego los datos son insertados en la base de ElasticSearch.

Se encontraron algunos desafíos a la hora de implementar la solución:

- Las diferentes fuentes de información tienen Ids distintos para sus publicaciones, el código de las publicaciones en mercadolibre tiene un prefijo "MLU" correspondiente a "Mercado Libre Uruguay", por esta razón estamos seguros que no va a haber colisiones en el futuro pues los códigos del gallito son sólo numéricos. Esto nos da la idea que para futuras fuentes de información se puede agregar prefijos artificialmente para evitar colisiones. Por ejemplo, para el gallito podría ser "GAL".
- Mantener la recolección de datos funcional va a llevar trabajo en el tiempo, a medida que se vayan introduciendo cambios en la forma que son presentados los datos por las fuentes.

Algunas otras decisiones de diseño que se tomaron fueron:

- Recordar la fuente del dato en un campo particular, con el objetivo de poder ser usado como filtro luego.

## Funcionalidades y Uso

Además de los filtros mencionados se incluyó una barra para buscar inmuebles por el título de la publicación.

A continuación se presentan capturas de pantalla de la aplicación web funcionando.

**Mi Nuevo Hogar** Search

**Venta/Alquiler**

- Venta 992
- Alquiler 458
- Alquiler temporal 171
- Alquiler temporario 142

**Sitio**

- MercadoLibre 929
- Gallito 834

**Moneda**

- UYU 1255
- USD 508

**Price Range**

\$10 - \$250000

**Ubicación**

- Piriápolis 319
- Pocitos 113
- Punta Colorada 103

1394 results found in 3ms

Ordenar por Precio (Menor a Mayor)

**Oferta Para 2 Personas**  
USD 45  
Apartamento en Piriápolis  
Alquiler temporal  
Dormitorios: 1  
Baños: 1  
Superficie: 200 Mts  
Publicado en MercadoLibre

**Monoambiente En Lo Mejor De ...**  
\$ 50  
Apartamento en Punta del Este  
Alquiler temporal  
Baños: 1  
Superficie: 35 Mts  
Publicado en MercadoLibre

**A Metros Del Mar, Cerca De La ...**  
USD 50  
Casa en Antoniopolis  
Alquiler temporal  
Dormitorios: 1  
Baños: 1  
Superficie: 50 Mts  
Publicado en MercadoLibre

**Apartamento En Lo Mejor De La...**  
USD 50  
Apartamento en Punta del Este  
Alquiler temporal  
Dormitorios: 1  
Baños: 1  
Superficie: 40 Mts  
Publicado en MercadoLibre

**Alquilo Casa En Arachania, La P...**  
USD 50  
Casa en Arachania  
Alquiler

**Alquilo En La Mejor Ubicación ...**  
USD 50  
Apartamento en Punta del Este  
Alquiler temporal

**Alquiler De Casas En Piriapolis**  
USD 50  
Casa en Piriápolis  
Alquiler temporal

**Alquiler Punta Del Diablo Piscin...**  
USD 50  
Casa en Punta del Diablo  
Alquiler temporal

Aplicando algunos filtros se tiene el siguiente resultado

**Mi Nuevo Hogar** Search

**Venta/Alquiler**

- Venta 992
- Alquiler 458
- Alquiler temporal 171
- Alquiler temporario 142

**Sitio**

- MercadoLibre 929
- Gallito 834

**Moneda**

- UYU 1255
- USD 508

**Price Range**

\$10 - \$250000

**Ubicación**

- Piriápolis 319
- Pocitos 113
- Punta Colorada 103

7 results found in 2ms

Ordenar por Precio (Menor a Mayor)

**Busco Terreno O Casa Vieja En B...**  
\$ 1111  
Terreno en Córdón  
Venta  
Superficie: 300 Mts  
Publicado en MercadoLibre

**Apartamento Centro Monoamb...**  
\$ 14000  
Apartamento en Centro  
Venta  
Baños: 1  
Superficie: 35 Mts  
Publicado en MercadoLibre

**Alquiler O Venta Casa A Estrena...**  
\$ 20000  
Casa en Parque del Plata  
Venta  
Dormitorios: 2  
Baños: 1  
Superficie: 600 Mts  
Publicado en MercadoLibre

**Las Piedras Vendo**  
\$ 65000  
Casa en Las Piedras  
Venta  
Dormitorios: 3  
Baños: 1  
Superficie: 400 Mts  
Publicado en MercadoLibre

**Apartamento 6to Piso Sobre La ...**  
\$ 95000  
Apartamento en Atlántida  
Venta

**Venta Apartamento Monoambi...**  
\$ 98000  
Apartamento en Centro  
Venta

**Hermoso Apto Frente Al Shopin...**  
\$ 100000  
Apartamento en Punta del Este  
Venta

## Evaluación y Resultados

Se obtuvo un sitio web de rápida respuesta y con una buena interfaz gráfica que además es responsiva. Las publicaciones son rastreables hasta su origen, en particular cada usuario puede ir a la publicación original clickeando en ella.

Existen inmuebles duplicados en nuestros datos, entre fuentes distintas e incluso dentro de la misma fuente se repiten publicaciones ya que los publicadores buscan tener el mayor alcance posible a potenciales consumidores. Una posible solución podría ser comparando direcciones, aunque no se llegó a implementar.

## Trabajo Futuro

- Agregar más fuentes de información, como por ejemplo InfoCasas.
- Automatizar la importación de nuevas publicaciones, de forma que se haga periódicamente y de forma incremental en el tiempo.
- Revisar validez de publicaciones creadas, se podría verificar periódicamente si una publicación está activa, ya sea por API o Scrapping, según la fuente.
- Realizar aplicación móvil nativa que brinde la funcionalidad del sitio web de mejor forma para dispositivos móviles.
- Agregar página en nuestro sitio web que muestre más detalles de cada publicación
- Extender a varios países, en MercadoLibre por ejemplo sería simplemente cambiar un parámetro en el llamado a la API para acceder a inmuebles de otros países. InfoCasas también tiene presencia en varios países latinoamericanos.

## Conclusiones

Nos pareció un trabajo desafiante y enriquecedor, el mayor desafío en nuestra opinión se dió en la consolidación de datos de las diferentes fuentes, y para el cual no siempre hay una única solución. Queda entonces a criterio de los desarrolladores determinar qué solución aporta más valor al usuario final en cada paso.

Sin duda creemos que un sitio así sería de gran interés para los usuarios finales, aunque probablemente no se alinee a los intereses de las empresas donde se publican los datos. Suponemos deben estar interesadas en mantener a los usuarios dentro de su propio ecosistema, en general se cobra por darle más prioridad en las búsquedas dentro del sitio por ejemplo.



# Referencias

- [1] API MercadoLibre - <https://developers.mercadolibre.com/api-docs/>
- [2] SDK Python MercadoLibre - <https://github.com/mercadolibre/python-sdk>
- [3] BeautifulSoup - <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- [4] ElasticSearch - <https://www.elastic.co/elasticsearch/>
- [5] ReactiveSearch - <https://opensource.appbase.io/reactivesearch/>
- [6] requests-html - <https://pypi.org/project/requests-html/>