

Recuperación de Información y recomendaciones en la web

Curso 2021

Grupo 02

Segunda entrega

Funciones de cine centralizadas

Autores

David Larrouy - 4.908.966-9

Enrique Castro - 5.373.074-9

Joaquin Bartaburu - 4.770.263-1

Augusto Alonso - 4.866351-3

Supervisores

Libertad Tansini

18 de Noviembre de 2021



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

Introducción

La pandemia de la COVID-19 ha tenido un gran impacto en la vida social de las personas. Si bien en el 2021 en Uruguay se están retomando distintas actividades como lo son el cine, teatro, bares, restaurantes, etc., nada es como era antes. Debido a las normas sanitarias, muchas de las actividades nombradas anteriormente tienen ciertas restricciones de horario, aforo, etc.

Los cines en particular, tienen un aforo mucho menor al que solían tener antes de la pandemia, y también hay menos funciones debido a que las salas deben ser desinfectadas entre las distintas funciones. Esto impacta directamente en la oferta disponible para los usuarios, ya que la cantidad de usuarios es prácticamente la misma, pero la oferta es mucho menor que antes.

En este trabajo, se presenta una herramienta cuyo objetivo es facilitar el acceso a entradas de cine para los usuarios mediante distintas técnicas de recopilación de información.

Problema

En Uruguay, cuando un usuario quiere ir a ver una película, primero debe elegir una cadena de cine a la cual desea ir. Existen varias en el país, como por ejemplo Movie, Grupocine, Life Cinemas, etc.

Debido a que hay una gran cantidad de cines de los cuales elegir, al momento de elegir una película, los usuarios generalmente visitan las distintas páginas web de estos cines para ver qué funciones hay y luego de ver las distintas opciones, se decantan por una de ellas.

El problema es bastante claro, no existe un lugar centralizado donde se puedan ver todas las funciones para todos los cines del país, sino que los usuarios deben visitar las distintas páginas web una por una antes de tomar una decisión sobre a qué función asistir.

Si bien el problema no surge a raíz de la pandemia de la COVID-19, la emergencia sanitaria empeoró el problema, ya que al haber menor cantidad de funciones, los usuarios deben visitar una mayor cantidad de páginas para ver las funciones disponibles ya que la posibilidad de encontrar una función en un día y hora que le sirvan al mismo, son menores.

Enfoque

El enfoque de la solución parte desde la perspectiva de lo que un usuario espera de un sistema que facilite la tarea de tener que decidir qué función de cine es la que mejor satisface la necesidad del usuario.

Para esto se pensó en una solución que como primer objetivo centralice la oferta de los diferentes cines disponibles para que el usuario no tenga que interactuar con diferentes páginas web por cada cine donde desee consultar información acerca de las funciones.

Esta oferta consistiría de las funciones de las películas disponibles en dichos cines, proporcionando información relevante de las mismas como cantidad de entradas disponibles, precios título, descripción y horarios.

Como segundo objetivo, ofrecerle al usuario mecanismos que faciliten el tomar una decisión de qué película ir a ver en función de los criterios de satisfacción para el usuario. Para este punto se pensó en utilizar filtros relevantes y acumulativos que restringen el conjunto inicial de películas ofrecidas en base a las preferencias del usuario.

Por último se armó un conjunto inicial de filtros en base a parámetros relevantes para elegir qué función se adapta mejor a las exigencias del usuario y se terminó seleccionando los dos encontrados más útiles, los cuales resultaron ser el cine donde se quiere ver determinada película y la fecha y hora de la función correspondiente.

Implementación

Arquitectura de la aplicación

La arquitectura de la aplicación se divide en cuatro componentes principales. A continuación se enumeran los mismos.

Base de datos

La base de datos almacena las películas y funciones recopiladas de ambos cines para luego ser obtenidas a través de la API de la aplicación. Para implementar la base de datos, se utiliza el manejador relacional PostgreSQL¹.

Worker

Es una aplicación hecha con Node.js² y su función principal es recopilar las distintas películas y sus respectivas funciones para los cines Movie y GrupoCine.

En el caso de Movie, el programa obtiene las películas y funciones mediante una api existente y provista por la misma cadena de Movie, por lo cuál es relativamente sencillo obtener dicha información. Simplemente se obtiene el listado de las películas disponibles y luego se itera sobre ellas para obtener las funciones para cada película.

Por otro lado, GrupoCine no cuenta con una API pública, sino que utiliza la técnica de server-side rendering, donde se obtienen los datos en el servidor y se provee el HTML con los datos ya poblados en el mismo³. Por esa razón, se utiliza Puppeteer, una librería de alto nivel para Node.js que permite controlar los motores de Chrome o Chromium para interactuar con páginas web de manera automatizada⁴. Para ello, primero se debe inspeccionar el código de la página de GrupoCine y diseñar una estrategia para obtener las funciones para una película utilizando selectores de HTML como *name*, *id*, *class*, etc. Esto permite simular la interacción del usuario con el sitio web y recopilar la información necesaria mediante el uso de dichos selectores.

Una vez obtenida las películas y funciones de Movie y GrupoCine, se realiza una correspondencia uno a uno entre ambas listas de películas con el fin de unificar las funciones de ambos cines para una película. Para ello, se utiliza la librería string-similarity, que dado dos strings retorna un grado de similaridad entre ambos strings. Esto permite comparar los nombres de las películas para ambos cines y, en caso de obtener un grado de similaridad mayor a un valor arbitrario, definir qué se trata de la misma película. Esto es necesario porque existen casos en donde los títulos de las películas son levemente distintos en ambos cines, como por ejemplo "007: Sin tiempo para morir" para el caso de Movie y "007 no time to die" en el caso de GrupoCine.

Finalmente, se borran los datos existentes en la base de datos con el fin de evitar inconsistencias, y se agregan las nuevas películas y funciones a la misma para luego ser consultadas por la API.

Idealmente, el programa ejecutado por el worker debería correr varias veces al días para eliminar las funciones con fecha de comienzo anterior a la fecha actual y también obtener nuevas funciones que pueden haber sido puestas a disposición luego de la última ejecución. Para los efectos de este trabajo, el worker se ejecuta manualmente.

API

La API es la interfaz entre la aplicación web y la base de datos. La misma fue implementada utilizando Node.js y el framework web Express⁵. La misma contiene los siguientes endpoints:

- **GET /movies**: retorna las películas disponibles para ambos cines.

¹ <https://www.postgresql.org/>

² <https://nodejs.org/>

³ <https://www.omnisci.com/technical-glossary/server-side-rendering>

⁴ <https://pptr.dev/>

⁵ <https://expressjs.com/>

- **GET /showtimes**: retorna las funciones disponibles para las películas de ambos cines. Provee la capacidad de filtrar las funciones por un rango con fecha de inicio y fecha de fin.

Sitio Web

Componente desarrollado en React⁶, cuenta con una UI sencilla que muestra al usuario de forma centralizada la oferta de películas de los cines anteriormente mencionados, a partir de invocaciones a la API. La web cuenta con una única página de index donde el usuario encuentra toda la oferta de películas en el período de dos semanas a partir de la fecha actual.

Las películas se muestran al usuario en forma de Cards, donde el usuario haciendo click en el botón Horarios puede ver la información sobre las distintas funciones ofrecidas en ambos cines para la película escogida. Las funciones le indican al usuario **dónde** (en qué sucursal del cine), **formato** (subtitulada, 2D, 3D, etc) y **horario** (fecha y hora de la presentación). Esta información es obtenida invocando el servicio /showtimes de la API.

Además se ofrecen 2 tipos de filtros, un filtro para escoger el cine del cual se quiere obtener su oferta (movie, grupo cine o ambos) y un filtro de fechas donde el usuario puede ingresar a través de un calendario y un conjunto de horarios preestablecido un rango horario en el cuál desea que hayan funciones en el o los cines, además pueden combinarse para restringir aún más la búsqueda.

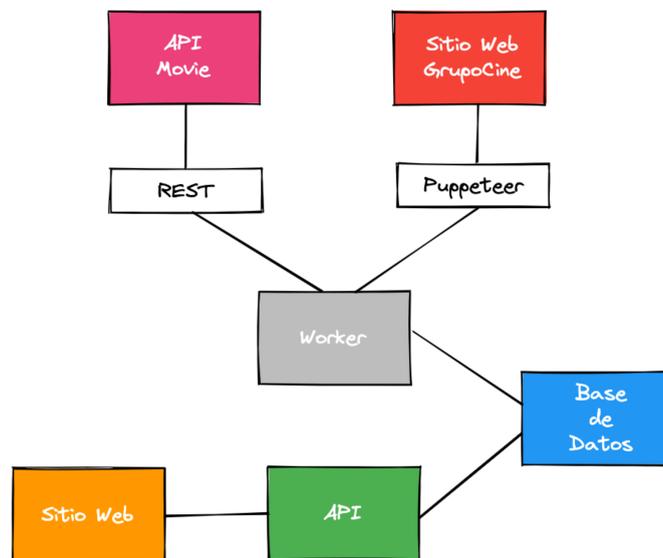


Figura x. Arquitectura de la aplicación

Funcionalidades y uso

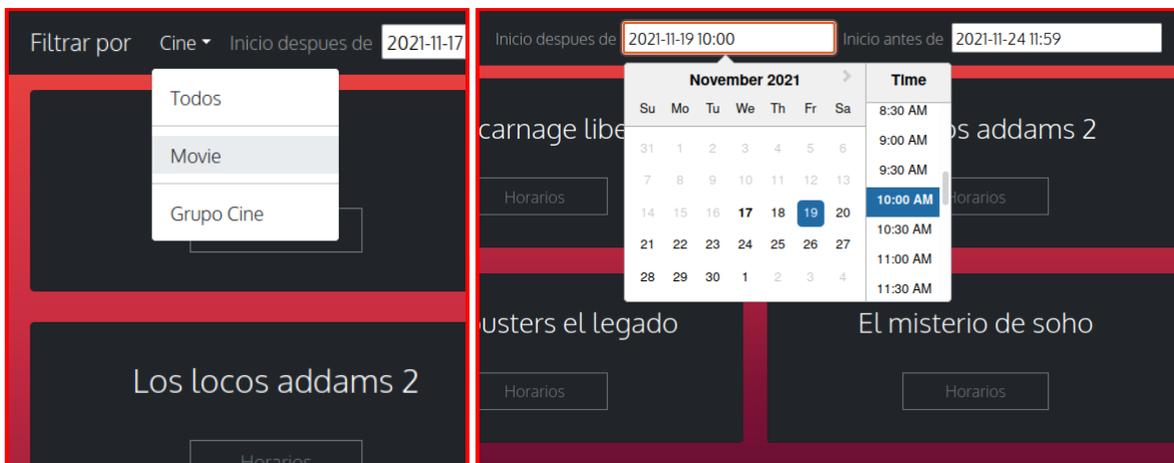
Se accede entonces a la web de la aplicación para ver la oferta disponible en los cines Movie y Grupo Cine

⁶ <https://reactjs.org/>

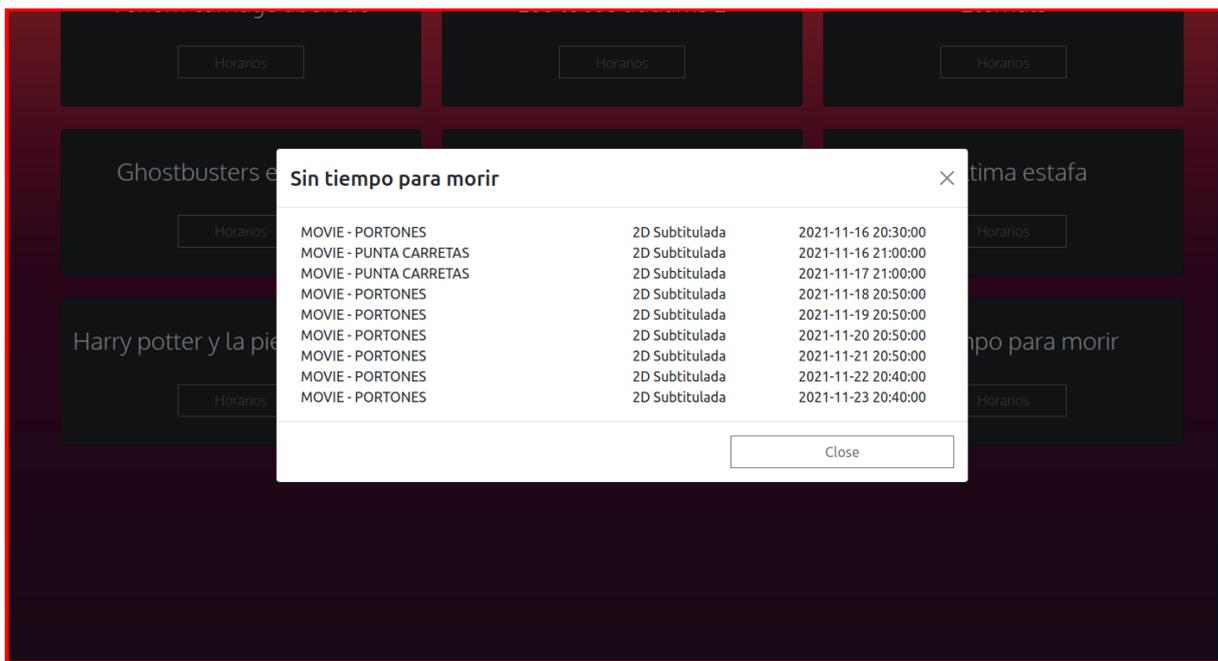
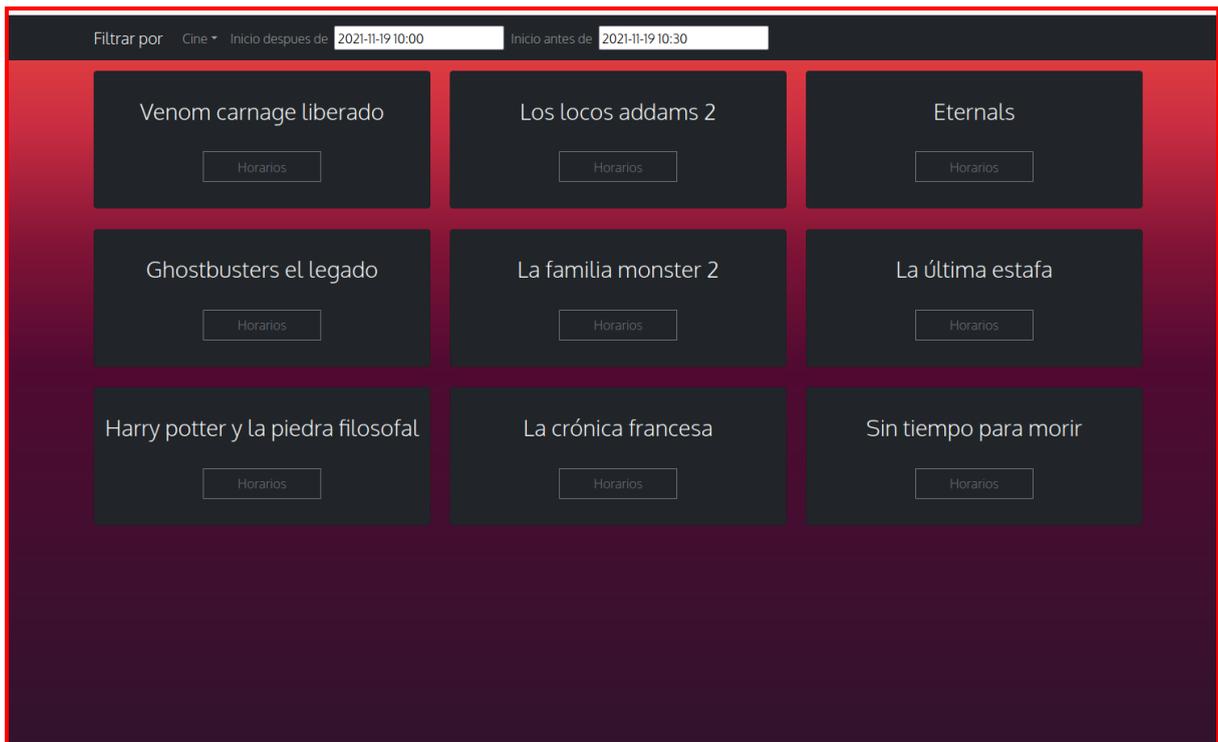


Podemos ver entonces los títulos de las películas que se ofrecen al usuario, con la opción de ver más detalles en cada una. Además se muestra en la barra arriba de la pantalla dos opciones de filtros para reducir la cantidad de películas a examinar.

Procedemos entonces a utilizar estos filtros para reducir nuestra oferta de películas. Elegimos un cine (e.j Movie) y además escogemos una fecha y hora de preferencia dentro de las fechas habilitadas (son las fechas para las cuales existe al menos una función ese día).



La aplicación de los filtros es inmediata al seleccionar los valores, como resultado podemos observar que la cantidad de películas se vio reducida en base a la aplicación de ambos filtros. Finalmente procedemos a seleccionar información adicional de uno de los títulos (e.j Sin tiempo para morir) y observamos el siguiente cuadro.



Nos encontramos entonces con una tabla con la información sobre la película “Sin tiempo para morir” que cumple con los filtros seleccionados anteriormente, es decir tiene al menos una función en el cine Movie a la hora seleccionada por el usuario. En este cuadro podemos apreciar el lugar dónde se presenta la función, el formato de la película y los horarios disponibles para ir a verla.

Evaluación y resultados

Se consiguió lograr un sitio web de una respuesta relativamente rápida, y donde la interfaz resulta sencilla de usar para el usuario. Debido a que no está cargada con contenido innecesario, decimos que es amigable para su uso y que un usuario puede lograr su objetivo sin dificultad ninguna. Los tiempos de respuesta son acordes a lo que esperábamos, debido a que el scrapping se realiza en tiempo real, pero no creemos que afecte para nada su funcionalidad.

En cuanto a resultados obtenidos, se puede ver en las capturas arriba que el funcionamiento del sitio web es el esperado. Al ingresar al sitio vemos las películas disponibles en los dos cines a los cuales consultamos, se despliega claramente el nombre de la película y damos la opción de apretar un botón para ver más detalles sobre los horarios de las funciones de la misma.

Una de las mayores dificultades encontradas fue al momento de agregar el filtrado por horario, debido a que esto requería un cambio en la arquitectura de la solución. Hasta ese momento no se utilizaba una base de datos, debido a que tanto la consulta a la API de Movie como el scrapping realizado a la página de GrupoCine se realizaban en tiempo real, y no había necesidad de guardar los datos. Para lograr que el filtrado efectivamente funcione, era necesario guardar los datos de las funciones en una base de datos, para luego una vez obtenidos los datos del horario introducidos por el usuario, consultar a la base y devolver aquellas funciones que se encuentren en dicho rango. Para esto hubo que involucrar el manejo de la base de datos, e introducir la lógica asociada a devolver los datos adecuados.

Conclusiones

Las impresiones del trabajo realizado fueron buenas. Se cumplió con el objetivo inicial que era centralizar las funciones de películas de distintos cines en un solo lugar.

Si bien existen muchas funcionalidades que hubieran aumentado el valor del trabajo como se explica en la siguiente sección, se entiende que esta primera versión es una buena aproximación a lo que se deseaba llegar desde un principio.

Por otro lado, una de las conclusiones a las que se llega es que en caso de que la aplicación fuera puesta en producción, se debería realizar un monitoreo de los sitios web que se tuvieron en cuenta, ya que, en caso de que alguno de ellos cambie, el funcionamiento de la aplicación se vería comprometido. Sin embargo, esto es una conclusión bastante lógica y que aplica a todos los sitios que utilizan técnicas de scrapping o utilizan APIs que cambian con el tiempo y cuya intención no fue el uso público de las mismas.

Otra conclusión que vale la pena mencionar es que el sistema en su conjunto presenta un stack variado de tecnologías para realizar las funcionalidades anteriormente mencionadas que lo hacen atractivo para desarrolladores que busquen expandir sus conocimientos y por lo tanto un proyecto muy interesante para trabajar.

Trabajo Futuro

Si bien GrupoCine y Movie son dos de los cines más importantes del país, incluir más cines en la aplicación aumentaría la oferta de películas en la aplicación y sería más fácil para un usuario encontrar una película. Además de esta razón, los cines mencionados solo tienen salas en Montevideo, por lo que la inclusión de otros cines podría aumentar la cantidad de usuarios de la aplicación considerablemente en caso de que la misma fuera puesta en producción.

Resulta muy cercana la posibilidad de expandir el dominio del sistema para proveer más información relacionada a la oferta de películas. De tal forma que el usuario pueda consultar otras características de la misma, en particular una imagen representativa de la película, con el objetivo de mejorar la UX en el sistema.

Actualmente, la aplicación solo permite filtrar funciones por horario o por película. Una futura iteración podría incluir más filtros, como por ejemplo, género, duración, idioma, etc. Esto permitiría reducir potencialmente el tiempo que los usuarios tardan en decidir qué película elegir.

Si bien la aplicación es a modo de demostración, en caso de que la aplicación sea puesta en marcha se debería utilizar algún servicio que ejecute frecuentemente el Worker que sincroniza las películas y funciones para que los datos obtenidos por la web sean más recientes.

Uno de los principales objetivos de este trabajo era tener la posibilidad de ver las entradas disponibles de una función, pero debido a diversos desafíos que no eran triviales, como por ejemplo la necesidad de utilizar tokens para obtener dicha información desde la página de GrupoCine, se decidió optar por no incluir dicha funcionalidad en la aplicación. Claramente, esta funcionalidad tiene un gran valor porque el hecho de que las películas aparezcan en el listado no significa que las mismas tengan disponibilidad de entradas.

Además de detallar si una función tiene entradas o disponibles o no, otra posible funcionalidad a implementar sería la de poder pagar por las entradas en la aplicación sin tener que redirigir al usuario al sitio web del cine donde se expedirá la entrada.

Finalmente, en una iteración futura se podría implementar la funcionalidad de registro de usuarios. Esto permitiría tener una base de datos de usuarios para poder realizar reportes sobre el uso de la aplicación, sugerir películas a los usuarios en base a otras películas que han visto, agregaría la posibilidad de puntuar películas, y muchas funcionalidades más.