

Recuperación de Información y Recomendaciones en la Web

Gabriel Sosa | 4892852
Gabriel Rodríguez | 4718068
Ana Laura Rodríguez | 4088377

Índice

Índice	1
Introducción	2
Descripción del problema	2
Enfoque de la solución	3
Diseño	4
Implementación	5
Funcionalidades y uso	6
Evaluación de resultados	7
Conclusiones	8
Trabajo a futuro	8
Referencias	10

Introducción

Diariamente se manifiesta un tráfico importante de personas que se trasladan desde el interior del país hacia Montevideo o viceversa; los fines de semana y en temporada de vacaciones o turismo, el volumen de pasajeros de líneas de transporte público interdepartamental se ha visto incrementado, siendo Tres Cruces una de las terminales con mayor afluencia de personas. Es así como surge la necesidad de una solución tecnológica que permita al usuario del transporte público saber en tiempo real, de forma rápida y sencilla: dónde y a qué hora pasa el ómnibus que debe abordar para llegar a su destino. Dada esta situación resulta de interés ofrecer una solución que permita al usuario, un acceso más ameno a la información, de forma que logre visualizar los horarios de los ómnibus de forma sencilla y ágil.

Descripción del problema

En la actualidad la información referente a los horarios y destinos de una línea de ómnibus se puede encontrar en el sitio web de cada compañía, ofreciendo al usuario poca flexibilidad sobre la información a la hora de realizar una búsqueda. Algunas compañías ponen a disposición una imagen con los horarios, mientras que, en el mejor de los casos se brinda una web con alguna opción de filtro de búsqueda, dejando al usuario una importante labor manual al momento de encontrar la opción más adecuada para realizar su viaje, y más aún si el usuario desea cruzar información de diferentes empresas o destinos.

Existen algunas soluciones como por ejemplo la página web de Tres Cruces ^[1] que si bien centraliza la información de diferentes compañías y destinos, solo permite ingresar búsquedas desde Montevideo o hacia Montevideo, no tiene filtros de búsqueda por horarios de salida y de llegada al destino, o por ruta. Por otra parte está Urubus^[2] que ofrece una web y una aplicación mobile. Además de centralizar la información de diferentes empresas de ómnibus, Urubus permite ordenar por horario de salida o llegada, duración del viaje y empresa, e incluye filtros útiles al usuario como ser el turno diario o la comodidad, pero que aún así es mejorable; por ejemplo, la web no permite realizar una búsqueda si no se indica fecha de retorno, lo que a priori el usuario puede no tener definido.

Enfoque de la solución

Se tomó la decisión de realizar una aplicación web, para brindarle al usuario la facilidad de poder utilizar el sistema sin tener ninguna aplicación instalada, de forma que pueda acceder mediante el pc o un smartphone. Para la implementación de la aplicación web se utilizó el framework Ruby on rails, dado que en el grupo de trabajo se contaba con conocimiento en la tecnología: es un framework pensado en el desarrollo web con infinidad de librerías/gemas y pensado para facilitar el desarrollo. Además se planificó desde un principio, utilizar las herramientas de scraping que brinda el propio framework, para obtener datos de la pagina de Tres Cruces ^[1] y tener una base para comenzar a trabajar.

Para poder persistir la información que el sistema necesita se requirió desarrollar un modelo de datos que permite almacenar la información de: las líneas de ómnibus, los recorridos, las compañías, los horarios, y las principales paradas de ómnibus por las que pasa cada línea. En la sección de diseño se especifica mediante un diagrama UML el modelo de datos que fue utilizado, en el cual se reflejan los aspectos antes mencionados. Se optó por un modelo de datos relacional, dado que la información que se desea almacenar es estructurada y con una construcción de índices adecuado podrá dar soporte a nuestra aplicación web sin mayores problemas.

Una vez definido el modelo de datos se trabajó en la extracción de los datos que alimentarían al sistema. Explorando el Catálogo de Datos Abiertos^[3] se descubrió un recurso con la información que se estaba buscando, y que presenta una forma más canónica que la extraída en primera instancia. Además especifica que corresponde al horario de buses confirmados para la temporada de invierno 2018. Se decidió trabajar con estos datos y no con los de Tres Cruces dado que estos últimos muestran aspectos de que están desactualizados.

Se utilizó la herramienta Spoon: un diseñador gráfico de transformaciones y trabajos asociado con el sistema de ETL(Extract Transform Load) Pentaho Data Integration o comunmente conocido como Kettle. Con Spoon se definieron las transformaciones para extraer los datos de un archivo^[4] en formato csv disponible en la web de datos abiertos del gobierno uruguayo. Con el fin de acotar el alcance del proyecto, solo se consideraron los datos de las líneas correspondientes a las compañías COT y COPSA ESTE. Tras la extracción de los datos para esas empresas, fueron almacenados de forma normalizada en nuestra base de datos relacional.

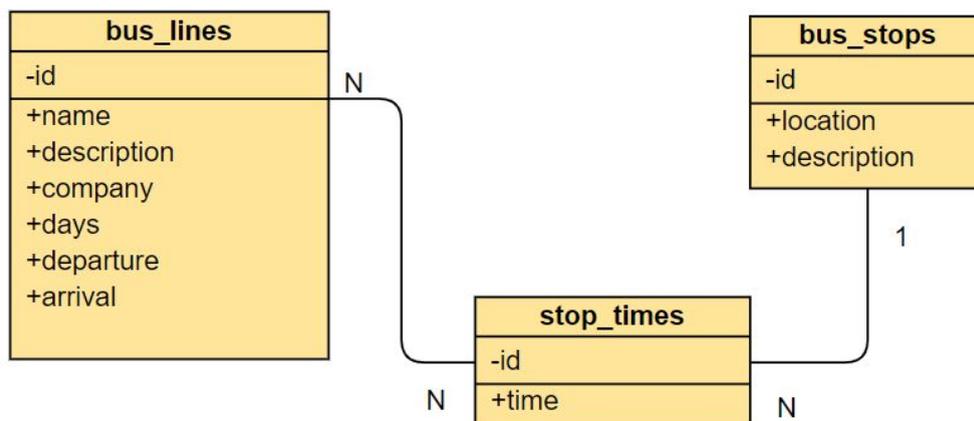
El prototipo a presentar cuenta con filtros: de origen de la ruta, destino de la ruta, empresa, horario de salida y horario de llegada. A su vez el modelo de datos contempla poder ubicar físicamente las paradas para poder dibujarlas en un mapa. Esta idea fue valorada inicialmente pero no se llevó a cabo la implementación pues no se encontraron datos que se pudieran importar a nuestro sistema

mediante scripting. Se alcanzó a contactar a 12 empresas vía e-mail, y solo una de ellas emitió una respuesta totalmente imprecisa sobre la ubicación de las paradas de sus líneas (solo nos dieron los nombres, ej. Plaza Cuba), con lo que obtener la ubicación demandaría un gran trabajo manual de buscar las paradas en google maps; obtener latitud, longitud e ingresar una por una las paradas de cada recorrido. Se consideró que a los efectos de esta prueba de concepto, no agrega valor significativo pero de igual manera lo mantuvimos en el modelo de datos puesto que del punto de vista “negocio” es una funcionalidad interesante.

Diseño

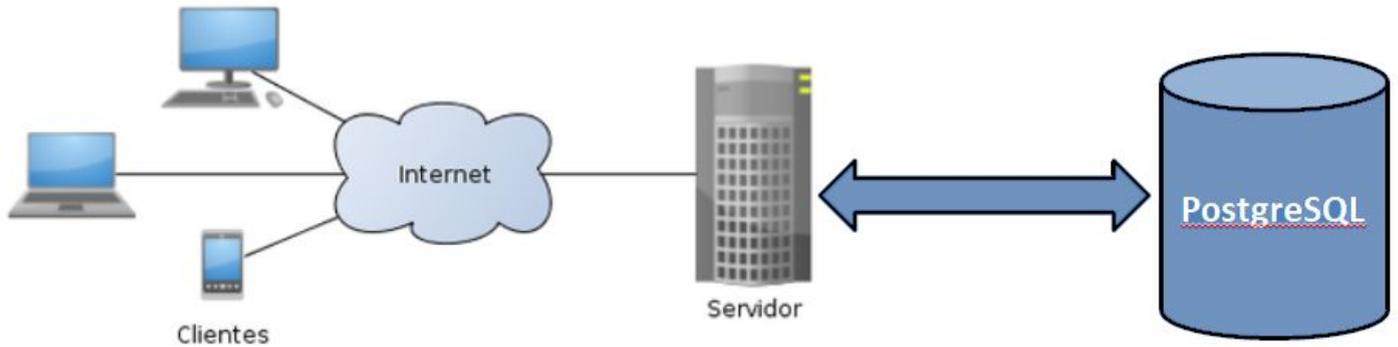
Modelo de datos:

En la siguiente imagen se puede ver el modelo de datos utilizado para almacenar la que obtuvimos tanto de la pagina de Tres Cruces ^[1] como de la web de datos abiertos del gobierno uruguayo^[3]. En esta estructura se almacena: los horarios de salida, horarios de llegada, origen, destino, compañía, días que la línea está activa (días de la semana y feriados).



Arquitectura:

La arquitectura tomada es la estándar: un modelo cliente servidor, se agrega una imagen a modo ilustrativo. De igual manera cabe destacar que Ruby on Rails permite escalabilidad horizontal en cluster.



Implementación

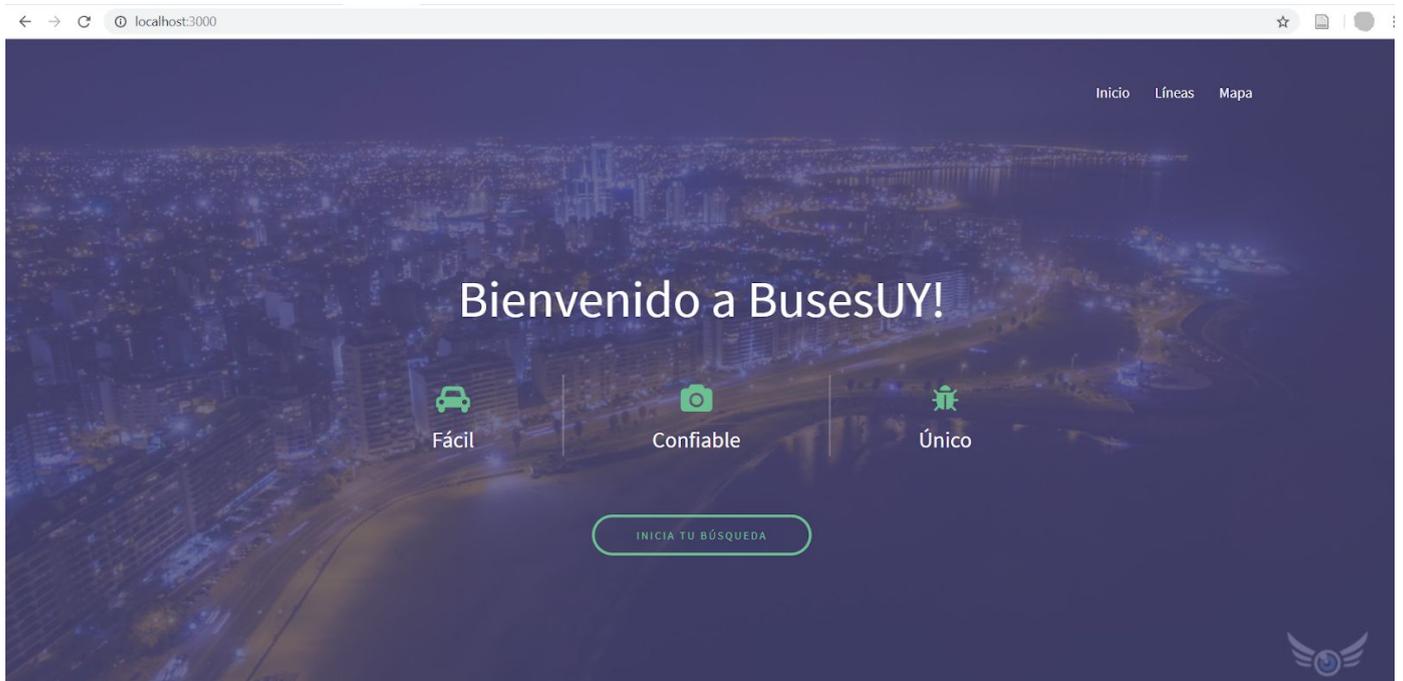
Ruby on rails se basa en el patrón de diseño Model View Controller y se implementaron las funcionalidades con el estándar del framework: para construir las vistas se utilizó HAML, los controladores con `*_controller.rb` y los modelos con `*_record.rb`.

Para la web se logró un diseño responsive que se adapta a diferentes tamaños de pantalla.

Mientras que para la base de datos se utilizó Postgres, porque se había pensado en utilizar el plugin PostGIS que permite calcular distancias entre coordenadas geográficas. Esto se iba a emplear para sugerir al usuario ir a una parada acorde a la menor distancia entre su ubicación y las paradas del recorrido del ómnibus que fuese a tomar. Finalmente no alcanzó a implementarse, pero quedó la decisión de diseño tomada para futuros cambios.

Funcionalidades y uso

Al ingresar el usuario al sitio web es redirigido a la página de bienvenida:



Luego se muestran las opciones para filtrar las líneas de buses por:

- 1- Origen
- 2- Destino
- 3- Compañía
- 4- Horario y día de salida
- 5- Horario de llegada

Cuenta además con un atajo rápido para el usuario que deseé salir ahora.

Una vez completadas las opciones aparecen en el cuadro de resultados las líneas que cumplen con las características indicadas, mostrando: nombre de la línea, empresa, hora estimada de salida y hora estimada de llegada. La información en el cuadro es presentada de forma que sea fácilmente legible para el usuario, y muestra un paginado en caso de que la cantidad de líneas obtenidas sea numeroso.

¿A dónde quieres ir?

ORIGEN

JAUREGUIBERRY

DESTINO

MALDONADO

EMPRESA

COPSA

QUIERO

Salir ahora Salir a las Llegar antes de las

13:00

DÍA

Sábado

BUSCAR

Encontramos los siguientes horarios para ti!

Línea	Empresa	Hora estimada	Llegada estimada
MONTEVIDEO - PUNTA DEL ESTE	COPSA	02:10	03:15
MONTEVIDEO - LAGUNA GARZON	COPSA	07:10	08:10
MONTEVIDEO - PUNTA DEL ESTE	COPSA	08:15	09:15
MONTEVIDEO - PUNTA DEL ESTE	COPSA	09:05	09:40
MONTEVIDEO - PUNTA DEL ESTE	COPSA	10:00	10:40
MONTEVIDEO - PUNTA DEL ESTE	COPSA	10:45	11:45
MONTEVIDEO - PUNTA DEL ESTE	COPSA	00:10	01:15

< Prev 1 Next >

Evaluación de resultados

Durante la ejecución de las pruebas, en primer lugar se consideraron opciones de búsqueda que no devolvieran datos, por ejemplo de líneas inexistentes, fuera de horario, o sin indicar destino y se controló que siempre se mostrará un mensaje adecuado al usuario.

¿A dónde quieres ir?

ORIGEN

AEROPUERTO INT. CARRASCO

DESTINO

LA PEDRERA

EMPRESA

COT

QUIERO

Salir ahora Salir a las Llegar antes de las

BUSCAR

Upps! No encontramos horarios para ti!

Luego, se tuvo en cuenta la consistencia de los datos ingresados para la búsqueda. Por ejemplo se controló no permitir ingresar el mismo origen y destino. Ese fue el único caso que se detectó en el que la entrada pudiera llegar a ser inconsistente, dado que los campos se presentan al usuario en forma de combo para que seleccione la opción que desea, evitando así la mayor cantidad de errores en el ingreso.

Finalmente se efectuaron consultas SQL directamente sobre la base de datos para verificar que los resultados obtenidos por la aplicación fueran los esperados.

Conclusiones

La información relacionada a las líneas de ómnibus interdepartamentales que hoy en día está disponible, requiere un gran trabajo manual por parte de los usuarios para la realización de búsquedas.

Por un lado están las compañías que publican los horarios y destinos, y no ofrecen nada más que eso, y por otro lado están los sitios de Tres Cruces o Urubús que si bien centralizan la información, no permiten realizar búsquedas demasiado personalizadas.

Después de haber analizado la realidad, consideramos que hay filtros de búsqueda que agregamos al prototipo que podrían ser de gran utilidad para el usuario, y no habían sido tenidos en cuenta; por ejemplo el filtro por horario de llegada y horario de salida. O el atajo de salir ahora, especialmente útil para el usuario que está apurado.

Otro punto importante es la presentación de los resultados al usuario. No resulta nada cómodo que el usuario tenga navegar por una infinidad de resultados hasta la línea que le resulta de interés. Desde este punto de vista, afirmamos que podría mejorarse mucho la calidad de los resultados, si se hiciera hincapié en la usabilidad.

Por otro lado el desarrollo de este prototipo implicó un amplio trabajo de investigación para la obtención de los datos. Consideramos que sería de gran utilidad contar con los datos geoespaciales de las paradas, para poder sacar cálculos de cercanía, con el fin de brindar mejores resultados.

Trabajo a futuro

Si bien originalmente se pensó en geolocalizar todas y cada una de las paradas de un recorrido para mostrarlo en un mapa, al no contar con datos que puedan cargarse de forma automática, fue una idea que quedó por el camino. Sin embargo consideramos que desde el punto de vista del usuario agrega mucho valor poder visualizar las diferentes rutas en un mapa; Un usuario podría querer transitar por una ruta costera para disfrutar del paisaje, por dar un ejemplo.

Otro punto pendiente sería el de incluir otras compañías y destinos. En una primera instancia se consideraron únicamente las líneas que realizan las empresas COT y Copsa Este. Agregar más datos no sería algo demasiado complicado, puesto que solo implicaría modificar las transformaciones realizadas en Spoon para la extracción de datos con Kettle, pudiéndose obtener la información de datos abiertos para las líneas y empresas del resto del país.

Otro aspecto que se podría explorar sería el de agregar rutas locales, por ejemplo los recorridos que hacen las líneas locales de Piriápolis y una vez que se tengan esas líneas, permitir combinaciones con transporte interdepartamental.

Resultaría interesante además agregar el origen y destino mediante un mapa, como hace la aplicación de ómnibus Montevideana “Como ir”^[5], y que en función de la distancia a las paradas el sistema sugiera hacer una u otra ruta y/o combinaciones.

Referencias

- [1] Web de Tres Cruces - salidas: <https://www.trescruces.com.uy/horarios-destinos/>
- [2] Portal de Urubus - <https://www.urubus.com.uy>
- [3] Catálogo de Datos Abiertos - <https://catalogodatos.gub.uy/>
- [4] Horarios ómnibus invierno 2018 - Datos abiertos
<https://catalogodatos.gub.uy/dataset/horarios-interdepartamentales-omnibus/resource/e62a3ef8-4e00-458b-bcbe-4b293ce0ebba>
- [5] Aplicación Como ir IMM - <http://www.montevideo.gub.uy/aplicacion/como-ir>
- Web de COT - salidas: <http://www.cot.com.uy/site/horarios.php>
- Web de COPSA ESTE - salidas: <http://www.copsa.com.uy/horarios-linea-este.jpg>