

# Sistema Informático para centralizar Eventos Running

Estefanía Della Mea  
4.517.469-2

Aldo Diaz  
5.031.609-7

Germán Larrosa  
4.561.973-5

## 1. INTRODUCCIÓN

El evento popularmente conocido como 'running' es la actividad de salir a correr al aire libre junto a muchas otras personas en un trayecto determinado.

El running es un hábito saludable que promueve la reducción del sedentarismo y concientiza la gente sobre la importancia de realizar actividad física regularmente para tratar los típicos problemas de la vida moderna como el estrés. Además que tiene una gran accesibilidad al no requerir grandes inversiones de dinero al iniciarse.

En el Uruguay, esta actividad es cada vez más común y se realiza casi semanalmente por parte de las grandes empresas con el fin de promover una marca.

## 2. DESCRIPCIÓN DEL PROBLEMA

Debido a la gran cantidad de Runnings que se llevan a cabo en nuestro país y a su vez la cantidad de entidades distintas que la promueven, se torna un problema a la hora de saber qué eventos y cuándo se realizan.

Hoy en día, existen webs que promueven los eventos de running, el problema es que hay carreras que están en ciertas webs y en otras no y vice-versa.

Por lo tanto, un usuario que desea planificar un fin de semana para participar en uno de estos eventos debe de navegar entre las distintas webs disponibles, las que conoce o encuentra mediante un buscador en los primeros resultados, viendo prácticamente la mayoría de eventos repetidos, o simplemente no tiene la opción de considerar un evento que está en un web que no conoce.

## 3. SOLUCIÓN PLANTEADA

Para la resolución del problema se plantea integrar y centralizar todas los eventos de running en un solo sitio. Con esto logramos que el usuario se libere de todos los problemas mencionados, que el evento de running sea aún más accesible y más adaptativo al tiempo de las personas debido a la completitud de la información disponibilizada.

La aplicación implementada lleva a cabo un proceso que va desde la información disponible en internet hasta la web donde se muestran los resultados. En un primer paso, se utiliza la herramienta Scrapy para la obtención de la información disponible en la web. Luego, esta información es volcada en una

base de datos luego de realizar un normalizado básico y limpieza de la información para facilitar la integración y lectura en etapas posteriores. Seguidamente, el backend es el encargado de tomar esta información y disponibilizar a través de servicios REST en la red. Finalmente, el frontend consume esta información para mostrarla de una manera amigable y prolija en un sitio web.

## 3.1 Arquitectura

En la figura 1 se muestra la arquitectura del software que implementa la solución planteada.

Por un lado, se tiene el módulo en python que utiliza la herramienta Scrapy para poder obtener la información de las webs que publican detalles de eventos de running que se van a realizar. La idea es tener una lista fija de estas webs e ir consultando cada cierto tiempo determinado por nuevos eventos o actualizaciones de los mismos.

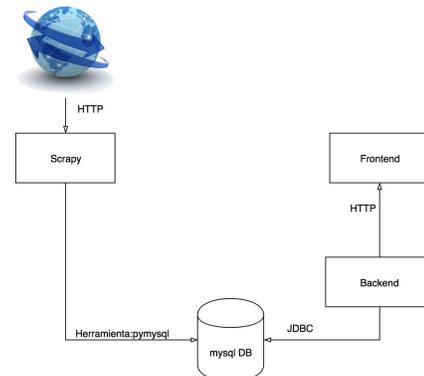


Figura 1 - Arquitectura del sistema

Luego, se tiene una base de datos MySQL para el almacenaje de la información. Dentro de la misma se define un esquema con una tabla que modela la información esencial del evento, como el título, kilómetros del recorrido, fecha, descripción y lugar.

El backend es el servidor que corre una aplicación en Java. Tiene el objetivo principal de exponer servicios con la funcionalidad de obtener la información de la base, procesarla y exponerla hacia el exterior.

El frontend es un servidor web que despliega una 'Single Page Application' en HTML5/CSS desarrollada utilizando el entorno de trabajo Angular. Se encarga de consumir la información del servidor y mostrarle al usuario de manera prolija y usable.

### 3.2 Implementación

En el entorno de Scrapy, existe una clase llamada Spider con las herramientas necesarias para hacer scrapping en los sitios webs. El término Spider se utiliza para denominar a aquellos módulos que hacen scrapping y derivan de la clase Spider. En la solución planteada se implementaron dos Spiders, una de ellas hace scrapping en el sitio “tucarrera.com.uy” mientras que la otra lo hace en el sitio “kronos.com.uy”. Ambos spiders trabajan sobre la estructura del DOM de las páginas que contienen la información sobre los runnings dentro del sitio. Por lo tanto, lo primero que se realiza es la obtención de la estructura del sitio, luego se toma la información deseada a través de las clases de los elementos. Por ejemplo, si las fechas se incluyen en una etiqueta HTML “h1” con el atributo “date-running”, en el Spider se busca un texto dentro de un elemento h1 con la clase “date-running”. Con toda la información transformada, se procede a descargarlo en la base de datos utilizando el paquete pymysql.

En el backend, se desarrolló una aplicación en la plataforma Java EE. Dentro del proyecto se definieron varios subproyectos encargados de cada uno de los procesos como muestra la figura 2. Esta modularización permite tener un proyecto apto para escalar y mantener. Por lo tanto, se define “controller” para contener a todos los módulos que exponen servicios, es decir, que representan a los nodos de comunicación; “dao” como contenedor de las colecciones de elementos (en este caso solo se trata de las carreras) y las operaciones que a éstos concierne; “entity” contiene a los modelos del sistema y finalmente “service” contiene a la implementación de los servicios junto a su interfaz que la enmascara.



Figura 2 - División del backend

La base de datos MySQL definida sólo contiene, para esta primera versión del sistema, una tabla con la información de los eventos de running. En la figura 3 se muestran los campos que se consideraron de importancia y que podían ser obtenidos de los sitios web son: título del evento, kilómetros del trayecto, fecha de realización, descripción del evento y el lugar donde se lleva a cabo.

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
TITLE	varchar(255)		YES	latin1	latin1_swedish...	select,insert,update,references	
KM	int(11)		YES	latin1	latin1_swedish...	select,insert,update,references	
DATE	datetime		YES	latin1	latin1_swedish...	select,insert,update,references	
DESCRIPTION	varchar(255)		YES	latin1	latin1_swedish...	select,insert,update,references	
PLACE	varchar(20)		YES	latin1	latin1_swedish...	select,insert,update,references	
ID	int(11)		NO			select,insert,update,references	auto_incre

Figura 3 - Modelo de Datos

Finalmente, el frontend es un servidor web que corre una “Single Page Application” desarrollada en HTML5/CSS en el entorno de trabajo Angular 4.0. Dentro de la misma se trabajó sobre un mismo componente debido a que su función actualmente es solo mostrar la información que recibe por parte del backend. Todas las operaciones que se utilizaron son nativas

al core de Angular, la visión del frontend es tener una aplicación lo más sencilla posible para asegurar una experiencia de usuario fluida y amigable.

### 3.3 Herramientas utilizadas

Tabla 1. Evaluación de Productos y Herramientas

Herramienta	Puntos Fuertes
Scrapy	Scrapping
MySQL	Open Source
Java EE	Veloz desarrollo
Angular	Ambiente ligero y de fácil arranque
Spring y sus componentes	Integración de herramientas esenciales en Java
pymysql	Conector de la base con los Spiders

## 4. PROBLEMAS ENCONTRADOS

Se encontraron diversos problemas tanto a la hora de implementar el sistema como al testearlo luego de finalizado. Varios de estos había ya sido identificados previamente, pero otros se encontraron en la marcha.

Uno de los principales problemas a tratar es la repetición de carreras entre los distintos sitios. Este problema es frecuentemente tratado en la Integración de Datos, se abordó con técnicas de identificación de la información. Para ello, se utilizó el algoritmo Levenshtein para obtener la distancia entre los títulos de las carreras debido que se observó que la misma carrera presentaba títulos distintos en distintos sitios, pero diferencias de escritura como “y” y “&”. Por lo tanto, todos las cadenas de caracteres que tengan cierta distancia hasta cierto umbral se las considera iguales. Para la obtención de este umbral se realizaron pruebas con textos similares encontrados.

Otro problema encontrado fue que en ninguno de los dos sitios especificaba la cantidad de kilómetros de las carreras. Por lo que tuvimos que parsear mediante la utilización de expresiones regulares el string del nombre en los Spiders, y ver en cuáles decía la cantidad de kilómetros en el nombre. Esto por supuesto no fue bueno, pues muchas carreras no tenían la cantidad de kilómetros en el nombre, por lo que tuvimos que poner un -1 en la tabla de la página web resultante. Otra desventaja de este método es que algunas carreras se pueden correr en diferentes formatos (en cuanto a los kilómetros), por lo tanto una carrera que se puede correr en modalidad 5k o 10k, la estamos limitando a una de ellas.

También se encontraron problemas respecto a la incompletitud de la información. Por ejemplo, en uno de los sitios se podía encontrar la localidad del evento, mientras que en otro se debía ingresar a otras urls para obtener la misma.

## 5. FUNCIONALIDADES

La utilización del sistema construido es al momento de querer ver cuáles son los eventos que hay disponibles para corredores y también cuáles va pasaron. Dado que nuestro sistema integra las carreras de dos sitios web, se muestran más eventos que si se mirara sólo en uno.

El sistema es perfectamente capaz de integrar en un futuro funcionalidades más intrínsecas a las carreras, por ejemplo la posibilidad de inscribirse a estas, el trayecto mostrado mediante un applet de Google Street View, puntos de referencia de interés, cómo llegar al punto inicial desde otro punto de la ciudad, entre varias otras funcionalidades que aportan un valor agregado al producto.

## 6. RESULTADOS

Se corrieron los módulos implementados para hacer scrapping en los sitios de Kronos y TuCarrera (Figura 4 y 5 respectivamente)



Figura 4 - Kronos.com.uy

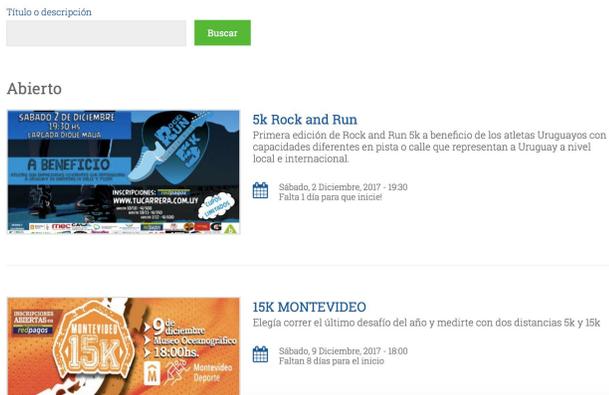


Figura 5 - tucarrera.com.uy

Esta información se guardó en la base de datos, obteniendo toda la información referente en una sola fuente como muestra la figura 6. En este caso, como el procesamiento de Integración se

realiza en el backend, la información puede estar duplicada o incorrecta (como los kilómetros -1 que se ven en la figura).

TITLE	KM	DATE	DESCRIPTION	PLACE	ID
5k Rock&Run	5	2017-12-02 00:00:00	-	Montevideo	1
AMEC	-1	2017-12-03 00:00:00	-	Montevideo	2
TRAVESIA MALVIN	-1	2017-12-09 00:00:00	-	Montevideo	3
15K Metropolitanos	15	2017-12-09 00:00:00	-	Montevideo	4
Trail Las Cañas	-1	2017-12-17 00:00:00	-	Las	5
CLUB REMEROS PAYSANDU	-1	2017-01-07 00:00:00	-	Paysandú	6
CLUB NAUTICO DE FLORES	-1	2017-01-13 00:00:00	-	Flores	7
BALNEARIO SAN LUIS	-1	2017-01-20 00:00:00	-	Cañeteles	8
CLUB DE PESCA PUNTA DEL ESTE	-1	2017-02-11 00:00:00	-	Maldonado	9
CLUB BANCO REPUBLICA	-1	2017-02-24 00:00:00	-	Cañeteles	10
5k Rock and Run	5	2017-12-01 19:30:00	Primera edición de Rock and Run 5k...		11
15K MONTEVIDEO	15	2017-12-01 18:00:00	Elegia correr el último desafío del añ...		12
San Felipe y Santiago 2017	-1	2017-11-01 19:15:00	San Felipe y Santiago, en su nueva f...		13
Club Nacional de Football 6.118km...	6	2017-11-01 10:30:00	El domingo 12 de noviembre a las 1...		14
Malvin 10kms y Sams 2017	10	2017-11-01 10:30:00	Sumate al desafío		15
Club Nacional de Football 6.118km...	6	2017-10-01 10:30:00	El domingo 15 de octubre a las 10:30...		16
7K BIVA 2017	7	2017-10-01 11:00:00	Comienza la temporada de carreras...		17
5k Peñarol 2017	5	2017-09-01 14:00:00	Como todos los años la 5K PEÑAROL...		18
Bimbo Global Energy 2017 Montevide...	-1	2017-09-01 09:00:00	El día 24 de setiembre comenzamos pas...		19
Super 1K	1	2017-09-01 15:00:00	Esta es una misión exclusiva para lo...		20
Half Maratón Montevideo 2017	-1	2017-09-01 10:00:00	YA TENEMOS DESAFIO para el 13/8...		21
MT10 Sierras de Rocha	-1	2017-08-01 10:30:00	MT10 no es una carrera más, es un p...		22
Club Nacional de Football 6.118km	6	2017-05-01 10:30:00	El 28 de mayo a las 10:30 hrs. la invit...		23
Maratón de Montevideo 2017	-1	2017-04-01 07:00:00	4ta edición de la media Maratón y Ma...		24
Costa Urbana Fit 2017	-1	2017-04-01 16:00:00	Llegó Costa Urbana Fit en su cuarta...		25
CORRE MONTEVIDEO EDICIÓN 2...	-1	2017-03-01 18:00:00	Corre Montevideo, ya marcada en el...		26

Figura 6 - Información obtenida de ambos sitios

Finalmente la información es procesada y desplegada en el front end como muestra la figura 7. La información que no se pudo obtener se colocó un guión como placeholder.

Eventos de running				
TRAVESIA MALVIN	-1	9/12/2017 00:00:00	-	Montevideo
15K Metropolitanos	15	9/12/2017 00:00:00	-	Montevideo
Trail Las Cañas	-1	17/12/2017 00:00:00	-	Las
CLUB REMEROS PAYSANDU	-1	7/1/2017 00:00:00	-	Paysandú
CLUB NAUTICO DE FLORES	-1	13/1/2017 00:00:00	-	Flores
BALNEARIO SAN LUIS	-1	20/1/2017 00:00:00	-	Cañeteles
CLUB DE PESCA PUNTA DEL ESTE	-1	11/2/2017 00:00:00	-	Maldonado
CLUB BANCO REPUBLICA	-1	24/2/2017 00:00:00	-	Cañeteles
15K MONTEVIDEO	15	1/12/2017 18:00:00	-	
Club Nacional de Football 6.118km San Ramon	6	1/11/2017 10:30:00	-	
Club Nacional de Football 6.118km San Ramon	6	1/11/2017 10:30:00	-	
7K BIVA 2017	7	1/10/2017 11:00:00	-	
Bimbo Global Energy 2017 Montevideo	-1	1/9/2017 09:00:00	-	
Half Maratón Montevideo 2017	-1	1/9/2017 10:00:00	-	

Figura 7 - FrontEnd con la información procesada

## 7. CONCLUSIONES Y TRABAJO A FUTURO

El trabajo actual en su primer versión de desarrollo logra poder realizar todo el ciclo completo partiendo desde más de un sitio web a un sitio particular, intentando resolver de la mejor manera los problemas más críticos del contexto como el de la integración de datos, obtención de la información, factibilidad del proyecto, entre otros. El usuario es capaz de visualizar todas las carreras ofrecidas por dos sitios en particular desde un solo sitio. Sin embargo, el sistema todavía tiene un largo trayecto por recorrer, por ejemplo, el diseño del sitio web puede ser mejorado y trabajado debidamente para ofrecer un aspecto llamativo y usable. También es posible aplicar herramientas de integración de datos más avanzadas para realizar pseudo-predicciones más exactas y poder mitigar los positivos falsos. Por último, se puede ampliar el modelo inicial de los datos obtenidos acerca de los eventos, por ejemplo, obteniendo imágenes característica del evento, link asociados con información más ampliada y más.

## **8. REFERENCIAS**

<http://www.tucarrera.com.uy/busqueda-publica>

<http://kronos.com.uy/?Q=calendario>

<https://doc.scrapy.org/en/latest/intro/tutorial.html>

<https://www.tutorialspoint.com/scrapy/>

<https://angular.io/docs>

<https://www.w3schools.com/angular/>

[https://es.wikipedia.org/wiki/Distancia\\_de\\_Levenshtein](https://es.wikipedia.org/wiki/Distancia_de_Levenshtein)