

Reconocimiento de emociones en Tweets

Grupo 18:

Guillermo Coduri - 4.583.667-0

Bruno Schiaffino - 4.625.835-0

Leonardo Oyharzabal - 4.479.515-4

Indice

Indice	2
Introducción	3
Twitter	3
#Hashtag y @Usuarios	3
Análisis de sentimiento	4
Problema	4
Enfoque de la solución	4
Arquitectura	5
Comunicación entre componentes	6
Frontend	6
Backend	6
Funcionalidades y uso	7
Conclusiones	7
Trabajo a futuro	8

Introducción

El presente informe tiene como objetivo presentar una aplicación elaborada con motivo del obligatorio de la asignatura Webir (Recuperación de información y recomendaciones en la web). La aplicación surge de la necesidad de analizar información disponible en algún entorno web y disponerla de forma ordenada al usuario final, de manera que a éste le sea de utilidad.

El documento se divide en varias secciones donde al principio se exponen algunos conceptos importantes para el entendimiento del mismo, como lo son la fuente de datos a analizar y el enfoque del análisis. Luego se da una descripción del problema atacado, o más bien, ejemplos de posibles problemas que existen que se podrían solucionar con la aplicación desarrollada. Más adelante, se detallan las tecnologías y herramientas utilizadas para alcanzar la solución y cómo se ensamblaron las mismas, para luego dar paso a las funcionalidades existentes el producto final. Finalmente se establecen las conclusiones del trabajo y puntos de mejora a futuro que podrían enriquecer el prototipo alcanzado.

Twitter

La fuente de información seleccionada en esta aplicación es la red social Twitter. Diariamente millones de personas, empresas, entidades gubernamentales y distintas organizaciones expresan sus pensamientos o simplemente realizan comunicados a través de esta red con los famosos Tweets.



Twitter cuenta con una masa de usuarios que se estima mayor a los 500 millones de usuarios, se generan aproximadamente 65 millones de Tweets por día y se manejan más de 800 mil peticiones de búsqueda diarias.

Por lo tanto, la cantidad de información es abismal, lo que provoca que analizar la misma pueda ser de gran utilidad para muchos factores. Dependiendo del contexto en el que se utilice esta información, puede ser importante tener que filtrar la misma, ya que al ser tan variados los generadores puede haber mucha que no sea de utilidad. Por otro lado, puede que mensajes con poco contenido informativo o hasta a veces poco éticos transmitan algo interesante.

#Hashtag y @Usuarios

En Twitter cada usuario puede ser referenciado con el símbolo @ y luego el nombre de usuario. Esto hace que este pueda identificar a otros usuarios en sus Tweets y a su vez lo puedan identificar a él también. Cada Tweet expuesto cuenta con el nombre del usuario que lo escribió a su lado. Además, existen los llamados “hashtags”, simbolizados con un # y el nombre o frase que los representa. Estos se asocian con eventos o cosas que son tendencia en alguna región o en el mundo entero, algún ejemplo puede ser #newiphoneX,

#uruguaymundial, #uruguaynoma o #eleccionesnacionales. Los “hashtags” son de especial interés porque tratan de temas de la actualidad al instante, los más referenciados son llamados “trending topics” los cuales son los más valiosos porque por algún motivo atrajeron a más gente.

Análisis de sentimiento

El enfoque que se tuvo para el análisis de la información en esta aplicación es la de detectar el sentimiento en los Tweets. De esta manera, un mensaje con buen contenido informático puede tener el mismo o hasta menor valor que uno más “vulgar”. Este tipo de análisis puede aplicarse a incontables ámbitos. Algunos de estos pueden ser la política, el deporte, la religión, etc.

Problema

Como se mencionó en la introducción del documento, el problema a resolver es el exceso de información y no disponer de una plataforma que revele algún valor en la misma.

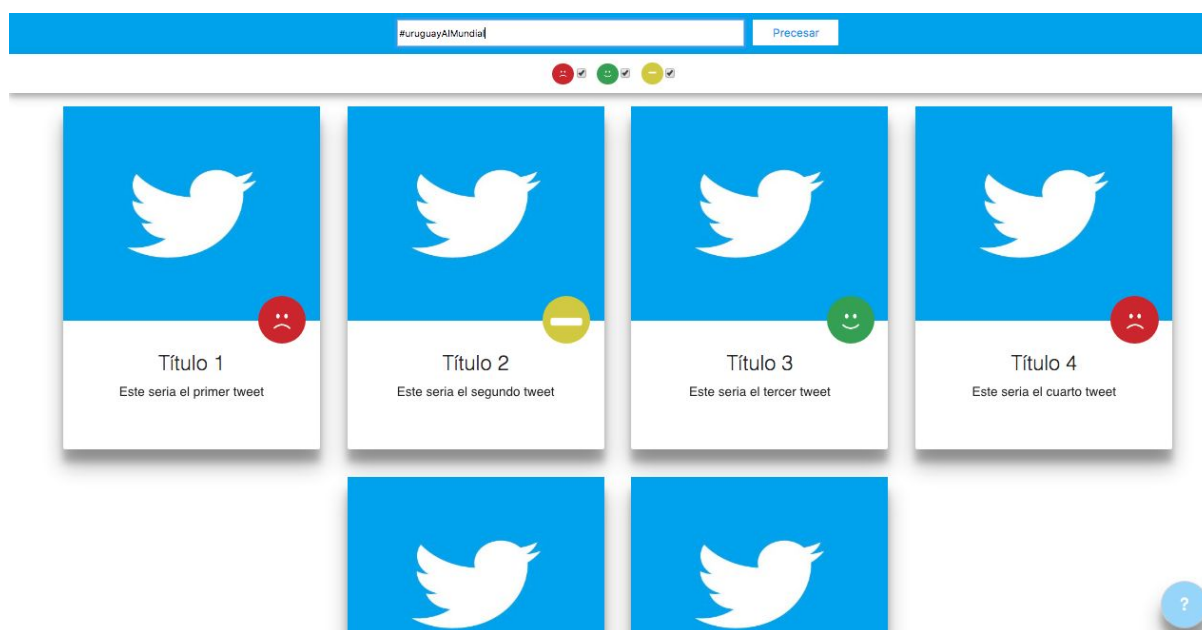
Entonces, en el caso especial de Twitter, se pueden encontrar muchos ejemplos donde hacer un balance entre pensamientos positivos o negativos sobre una temática puede ser interesante, se ven a continuación algunos de estos ejemplos:

- Políticas: En unas elecciones saber las connotaciones sentimentales de las opiniones de los votantes sobre un candidato pueden ser de gran validez como para elaborar una encuesta. Lo mismo sobre la popularidad de un presidente (los cuales es muy común que hoy en día tengan una cuenta de Twitter y la utilicen), reacciones ante leyes aprobadas, rendiciones de cuentas, presupuestos, etc.
- Empresas: Una empresa lanza un nuevo producto, ejemplo, el nuevo Iphone X de Apple y la gente opina sobre el mismo, se puede sacar buena información sobre la reputación del dispositivo a través de los Tweets.
- Deportes: Reacciones ante acontecimientos en eventos deportivos. Copas del mundo, olimpiadas, análisis sobre resultados de encuentros, el pensamiento de la gente ante escándalos como el pasado en la FIFA el año pasado o casos de dopaje en competiciones de atletismo.
- Noticias: Debe ser hoy en día Twitter una de las mayores herramientas para trasladar noticias. Noticias como desastres naturales, ataques terroristas (que lamentablemente son noticia bastante corriente en estos días) generan conmoción global y generan una gran cantidad de información analizable.
- Espectáculos: Además de para fines informativos, hay mucha gente que disfruta con información sobre el ambiente del espectáculo, el cine o concursos del estilo reality show. Estudiar el por qué, qué tipos de realities y la reacción de los usuarios sobre estas temáticas puede ser de interés.

Enfoque de la solución

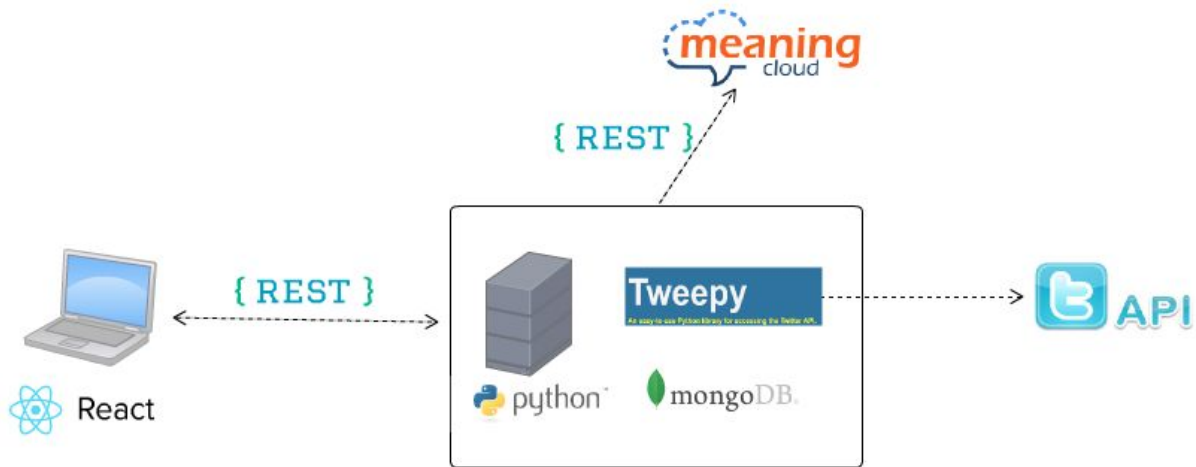
La solución propuesta es una aplicación cuyo objetivo es facilitar la obtención del análisis de sentimientos de los datos de los Tweets a través de una interfaz web. Se realiza un backend que obtiene los Tweets de una API disponibilizada por Twitter y los envía a otra que realiza el análisis de sentimientos. A su vez, la información de estos tweets se almacena junto su análisis en una base de datos no relacional. Si bien no es indispensable almacenar dichos tweets para lograr visualizarlos en el frontend, se entiende que podría llegar a ser útil en un futuro contar con una extensa base de datos de tweets junto con sus respectivos sentimientos. Por ejemplo, se podrían aplicar algoritmos de machine learning para detectar patrones, los cuales pueden ser utilizados para determinar de antemano el posible impacto de un posible tweet.

A continuación podemos ver una imagen de la interfaz web implementada



Arquitectura

En cuanto a la arquitectura del sistema, se optó por utilizar una arquitectura del tipo cliente-servidor. El cliente consta de una interfaz web, mientras que en el servidor se hostea todo lo relacionado a la lógica de negocios. Como podemos ver en el siguiente diagrama, se utiliza el plugin Tweepy el cual interactúa con la api de Twitter para obtener información de twitters. Adicionalmente, podemos ver que el backend interactúa con un sistema externo como Meaning cloud a través de REST para obtener el sentimiento del tweet asociado. A su vez, se tiene una instancia de base de datos de MongoDB para persistir la información obtenida.



Comunicación entre componentes

Para la comunicación entre el frontend y el backend, se expone una API Rest del lado del backend, se utiliza el estilo de arquitectura REST para el intercambio de datos entre los componentes. Esto brinda la posibilidad de desacoplar completamente los mismos, favoreciendo la mantenibilidad del sistema. Otra de las ventajas de utilizar este estilo de arquitectura se puede ver si se desea agregar un cliente mobile. Este último, podría interactuar con la API Rest y de este modo no tener que re implementar la lógica de negocio.

Frontend

El mismo se encarga de la interacción entre el usuario final y el sistema. Para esta interacción se implementó una interfaz web en la cual se listan los Tweets junto con un título y una pequeña descripción. A su vez, se indica de forma gráfica si se trata de un tweet en el cual se detecta algún sentimiento positivo, negativo, o no se detectan sentimientos. Adicionalmente, se cuenta con un header mediante el cual se pueden filtrar los Tweets listados según su sentimientos asociado, junto con un buscador para listar tweets según hashtag.

En cuanto a tecnologías, la interfaz web fue implementado con React JS, una moderna librería de JavaScript la cual es utilizada por grandes aplicaciones como Facebook y Netflix. A su vez, para la implementación de la parte visual, se tomaron como referencia patrones de diseños de Material Design para lograr un producto agradable estéticamente.

Backend

El mismo se encarga de obtener los tweets mediante peticiones REST a la api de Twitter para luego guardarlos en la base. Se utiliza Python para desarrollar este componente.

El sistema de base de datos seleccionado fue Mongo DB, el cual es no relacional (NoSql) orientado a documentos y está desarrollado bajo el concepto de código abierto. En lugar de guardar los datos en tablas como se hace en las base de datos relacionales, MongoDB guarda estructuras de datos en documentos similares a JSON con un esquema dinámico, haciendo que la integración de los datos en ciertas aplicaciones sea más fácil y rápida.

Cada vez que se genera un request se obtienen los Tweets relacionados al usuario o hashtag, se realiza el análisis de sentimientos y se persiste toda esa información en la base. Además se le devuelve al request la información resultante del análisis.

Funcionalidades y uso

La aplicación se trata básicamente de un buscador. En este buscador el usuario puede ingresar tanto un “hashtag” como un nombre de usuario. Esa búsqueda obtiene los Tweets referidos al hashtag o usuario y realiza el análisis de sentimiento sobre los mismos. Una vez procesados se listan los resultado del análisis.

Filtros

Mediante la interfaz web, se podrán filtrar los tweets analizados según sentimientos. Para esto se despliegan en el menú tres checkboxes los cuales se utilizan para cumplir esta funcionalidad.

Conclusiones

En lo que al producto alcanzado se refiere, se entiende que es un prototipo básico pero que cuenta con una funcionalidad interesante. El explorar la información de una red social adquiriendo un componente emocional es algo novedoso y que puede ser utilizado para hacer análisis en temáticas críticas a nivel mundial.

En cuanto a lo referido a la temática de la asignatura, se entiende que se elaboró una solución acorde a los conceptos de recuperación de información en la web, ya que se está analizando una masa de datos y la misma es expuesta de una manera diferente, procesada y analizada para que el usuario pueda sacar sus propias conclusiones.

Otro aspecto importante fue la elección de las tecnologías a utilizar, las cuales son de uso frecuente en la actualidad y generaron un desafío extra ya que se debió investigar sobre las mismas para llegar a una correcta implementación. En este sentido, se valora mucho la utilización de React para el frontend y el uso de Mongo DB para la persistencia de los datos.

En general, fue un trabajo interesante de llevar a cabo ya que el fin del mismo (recuperar información existente en la web y presentarla) no es el tipo de proyecto a los que los integrantes del grupo está acostumbrado. Este fue punto que llevó a la reflexión de por qué no existen más propuestas del estilo en la Facultad de Ingeniería, siendo que el análisis de datos es una herramienta potente en un mundo globalizado y con tanta información

disponible. Todo lo anterior, sumado a las tecnologías aprendidas y los recursos utilizados, permiten concluir que fue una buena experiencia de trabajo.

Trabajo a futuro

Como trabajo a futuro, se puede explotar la herramienta desarrollada en muchos ámbitos, en la sección relacionada al problema a resolver se expresaron varias de ellas. Esto implica desarrollar análisis más detallados sobre los Tweets que solamente un cuantificador de expresiones positivas o negativas. Además la capacidad de agregar más filtros y obtención de reportes más ricos serían buenas funcionalidades a agregar.

Este tipo de aplicaciones con análisis de sentimientos podría verse beneficiado con un desarrollo en el área de Machine Learning, teniendo la posibilidad de generar predicciones en base a los comportamientos de las personas y de sus expresiones. Este es un paradigma que ha tomado gran fuerza en estos días y definitivamente agregaría muchísimo valor al producto.