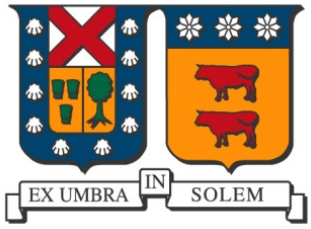
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
CIUDAD - CHILE



“DISEÑO DE UNA GUÍA DE CONDUCTAS PARA LA VIRALIZACIÓN DE ALTO IMPACTO DE CONTENIDOS EN TWITTER.”

CARLOS ALBERTO ANDRADE CABELLO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: José Luis Martí Lara
Profesor Correferente: ???

Diciembre - 2018

DEDICATORIA

Considerando lo importancia de este trabajo para los alumnos, este apartado es para que el autor entregue palabras personales para dedicar este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

AGRADECIMIENTOS

Considerando la importancia de este trabajo para los alumnos, este apartado se podrá incluir en el caso de que el autor desee agradecer a las personas que facilitaron alguna ayuda relevante en su trabajo para la realización de este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

RESUMEN

Resumen— El resumen y las palabras clave no deben superar la mitad de la página, donde debe precisarse brevemente: 1) lo que el autor ha hecho, 2) cómo lo hizo (sólo si es importante detallarlo), 3) los resultados principales, 4) la relevancia de los resultados. El resumen es una representación abreviada, pero comprensiva de la memoria y debe informar sobre el objetivo, la metodología y los resultados del trabajo realizado.

Palabras Clave— Cinco es el máximo de palabras clave para describir los temas tratados en la memoria, ponerlas separadas por punto y comas.

ABSTRACT

Abstract— Corresponde a la traducción al idioma inglés del Resumen anterior. Sujeto a la misma regla de extensión del Resumen.

Keywords— Corresponde a la traducción al idioma inglés de Palabras Clave anteriores.

GLOSARIO

Aquí se deben colocar las siglas mencionadas en el trabajo y su explicación, por orden alfabético. Por ejemplo:

Red Social: Sitio Web, aplicación o cualquier plataforma que utilice internet para conectar a personas a través de diversas dinámicas, como lo son el compartir y consumir contenido generado por otros usuarios, mensajería, etc.

Twitter: Red social de Microblogging en la cual los usuarios registrados comparten contenidos llamados Tweets.

Tweet: Contenido generado por lo usuarios de la plataforma Twitter, se basa en un texto limitado a 280 caracteres. Estos pueden poseer imágenes, videos, enlaces, entre otros.

Thread (Hilo): Funcionalidad de la plataforma Twitter, la cual permite escribir varios Tweets que se referenciarán de forma secuencial, permitiendo visualizar contenido extenso.

Cronología: Cuando un usuario sigue a otro, este se suscribe al contenido generado de la persona a seguir. Los tweets se presentan en la página inicial de twitter de cada usuario, a esta colección ordenada de forma cronológica se llama Cronología o timeline (TL) en inglés.

Retweet: Abreviado como RT, es la forma de compartir contenido generado por otro usuario de la plataforma, al hacer RT, los seguidores de la persona que realiza el retweet verán en su cronología el tweet original.

Favorito (Twitter): Tambien llamado Like o ”Me Gusta.es una funcionalidad presente en cada tweet, en el cual el usuario puede demostrar afinidad con el contenido presentado.

Respuesta (Twitter): Es un tweet escrito a modo de respuesta a otro tweet, generando dialogo entre los participantes.

Seguidor: Cuenta suscrita a los contenidos de otra, también es llamado Follower, por su nombre en inglés.

Seguido: Cuenta a la cual se encuentra suscrita otra cuenta.

Viralización: Fenómeno en el cual los contenidos se propagan de forma rápida e independiente, sin publicidad ni marketing, llegando de forma exponencial a nuevas personas.

Contenido Viral: Unidad de información que se ha propagado de forma rápida a una cantidad muy grande de usuarios.

Tópico: Tema (Idea o categoría) de una palabra o documento.

ÍNDICE DE CONTENIDOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

INTRODUCCIÓN

Debe proporcionar a un lector los antecedentes suficientes para poder contextualizar en general la situación tratada, a través de una descripción breve del área de trabajo y del tema particular abordado, siendo bueno especificar la naturaleza y alcance del problema; así como describir el tipo de propuesta de solución que se realiza, esbozar la metodología a ser empleada e introducir a la estructura del documento mismo de la memoria.

En el fondo, que el lector al leer la Introducción pueda tener una síntesis de cómo fue desarrollada la memoria, a diferencia del Resumen dónde se explicita más qué se hizo, no cómo se hizo.

CAPÍTULO 1
DEFINICIÓN DEL PROBLEMA

En la actualidad la sociedad chilena se encuentra hiperconectada, con una penetración del 71,7 % de personas con conexión a internet, de las cuales, el 94 % [1] se conecta principalmente a través de un dispositivo móvil. Estas personas utilizan en promedio 17 aplicaciones, siendo estas en su mayoría redes sociales. Lo anterior convierte a las redes sociales en uno de los mercados más interesantes para difundir contenidos debido al alcance y penetración que presentan.

Una de las redes sociales más utilizadas es la plataforma de microblogging Twitter, en la cual se pueden compartir diversos contenidos, en una publicación llamada tweet, la cual posee un límite de 280 caracteres (anteriormente 140).

Recientemente fue estrenada una funcionalidad en twitter llamada ”Threads” o Hilos, en la cual es posible escribir diversos tweets seguidos, los cuales se verán agrupados. Esta funcionalidad es muy útil cuando la información que se desea compartir es mucho más extensa como para ser presentada en un único tweet.

Debido a esta nueva funcionalidad presentada, es posible difundir contenidos que anteriormente eran más difíciles de viralizar debido a su extensión, lo que ha generado un gran interés para diversas entidades que buscan expandir sus receptores de información (ya sea clientes, adherentes políticos u otros, según sea quien difunda el contenido).

Considerando lo interesante del mercado presentado, en conjunto con las nuevas funcionalidades introducidas en las diversas plataformas sociales, es que empresas, organizaciones, entidades de gobierno e incluso personalidades públicas están invirtiendo en personal (Community Manager) para posicionar entre los usuarios su marca o contenido asociado, los cuales actualmente puede que posea o no los conocimientos necesarios para la difusión de contenidos.

Al existir dineros invertidos, es de gran interés poseer conocimientos específicos para generar un mayor impacto y viralización de los contenidos desarrollados, por lo que se reconoce como problema el no existir una guía o manual de buenas prácticas para difundir contenido a través de twitter.

Finalmente, se reconoce que una guía de buenas prácticas para la viralización de contenidos aporta un gran valor para las diversas entidades que desean posicionar productos, marcas o campañas en la plataforma social, puesto que optimizaría los recursos invertidos para este objetivo.

[1] Chile lidera la penetración de internet en la región y el smartphone continúa siendo el favorito Fuente: Emol.com - https://www.emol.com/noticias/Tecnologia/2017/05/04/856853/Chile-lidera-la-penetracion-de-internet-en-la-region-y-el-smartphone-continua-siendo-el-favorito.html

1.1. Identificación del problema

Se identifica como problema la dificultad de popularizar contenidos en una plataforma tan utilizada y llena de contenido como lo es Twitter, por lo que se busca analizar una gran cantidad de Hilos de tweets de diferente largo, para así poder obtener relaciones que puedan llegar a indicar factores de éxito en la difusión y alcance de estos.

De manera inicial se posee un dataset de 503 threads, los cuales en conjunto suman 8,894 tweets.

Cada registro posee id, número del thread perteneciente, timestamp, contenido, cantidad de retweets, cantidad de likes y cantidad de replies, tal como se puede apreciar en la figura 1.

	A	B	C	D	E	F	G
1	id	thread_number	timestamp	text	retweets	likes	replies
2	999307110902050818	Thread 1	1527088356	Extraordinary evidence at Treasury committee from Jon Thompson, CEO of HMRC on customs and Brexit today https://t.co/DJhIQhmVwJ	66	59	5
3	999307395712143360	Thread 1	1527088424	The Brexiter favourite Max Fac - would cost business between £17 and £20bn a year - that's almost 1% of GDP - jus&#65533;?&#65533; https://t.co/0MwIcwre4t	83	107	10

Figura 1: Ejemplo de registros de los datos inicialmente obtenidos.
Fuente de los datos: Kaggle.

Los registros están divididos en 5 archivos, separados por rangos de cantidad de tweets en threads, siendo estos largo de 5 a 10, 10 a 15, 15 a 20, 20 a 25 y 25 a 30 tweets.

Además de la información que ya se posee de manera inicial, es necesario rescatar información de Twitter a través de la API, atributos del usuario como lo son el id, cantidad de seguidores, seguidos y cantidad de tweets, los cuales son necesarios para el análisis a realizar.

Es necesario almacenar toda la información obtenida de forma que sea posible manipular los datos de forma rápida a través de código y herramientas de análisis, por lo que se debe definir una estructura y forma de almacenamiento para los datos.

1.2. Objetivos

El objetivo general de esta memoria consiste en Diseñar y validar una guía de conductas para la viralización de alto impacto de contenidos en la plataforma Twitter a través de la funcionalidad de threads, mediante el análisis de múltiples dimensiones de diversos threads.

Objetivos específicos

- Estudiar el impacto generado según el largo, la emoción y el tópico de los threads, para identificar factores en común en los mensajes populares.
- Estudiar la relación entre el autor y el impacto de los threads, para reconocer la importancia del autor en relación al impacto del contenido.
- Analizar la relación de Favoritos, RTs y respuestas del tweet principal y los siguientes en un thread, para estudiar el comportamiento viral de estos.
- Examinar y asociar relaciones descubiertas entre variables de estudio, para proponer buenas prácticas de difusión de contenido.

1.3. Alcances

Para el trabajo a realizar se analizarán los atributos de al rededor de 9,000 tweets, agrupados en aproximadamente 500 threads. Estos documentos de estudio fueron obtenidos de la plataforma Kaggle [2]. Los documentos de estudio se encuentran escritos en inglés, algunos sólo poseen texto, sólo elementos multimedia, o ambos, por lo que se deberá realizar una limpieza inicial.

De manera inicial se poseen cinco archivos de valores separados por coma (.csv), los cuales tienen los siguientes atributos: id, número del thread perteneciente, timestamp, contenido, cantidad de retweets, cantidad de likes y cantidad de replies.

Se busca realizar un análisis tanto del impacto del contenido, como el impacto del usuario en sí en la viralización del thread, por lo que se debe rescatar la información del usuario creador.

[2]Twitter Threads: https://www.kaggle.com/danielgrijalvas/twitter-threads/

CAPÍTULO 2
MARCO CONCEPTUAL

A continuación se presenta el marco conceptual, que sirve como base para el trabajo a desarrollar. Se reconoce la plataforma, las técnicas de almacenamiento y los conocimientos a aplicar en la solución propuesta.

2.1. Plataforma Twitter

Twitter es una plataforma de microblogging desarrollada por Jack Dorsey el 2006, en la cual es posible compartir contenido, ya sea texto, videos, imagenes, entre otros, los cuales están limitados a un largo de 280 caracteres. Los usuarios utilizan la plataforma para compartir diversos tipos de contenido, como sus pensamientos a lo largo del día, noticias que encuentren interesantes, comentar programas de televisión, política, entre otros.

Figura 2: A la izquierda: Ejemplo de un TL, A la derecha: Ejemplo de un perfil público
Fuente: Twitter para Android.

Cada usuario genera contenido en su propia cuenta, por lo que para consumir el contenido de otra cuenta, los usuarios deben seguir a esta, lo que genera que su contenido aparezca en el inicio de Twitter. A esto se le llama cronología o Timeline (TL), un ejemplo de esto aparece en la imagen izquierda de la figura 2. En esta vista, los usuarios pueden ver los contenido generados por las cuentas que sigue el usuario.

Hay que dejar en claro que la relación entre seguidores no es simétrica, puesto que al momento de seguir a alguien, la relación es dirigida, por lo que se crea la distinción entre seguidores y Seguidos.

Una cuenta puede seguir a muchas personas y tener muy pocos seguidores, lo que genera que éste consuma contenido pero su contenido no sea visualizado. El caso contrario es cuando una cuenta posee muchos seguidores pero muy pocos seguidos, es decir, una cuenta popular entre usuarios. En la imagen de la derecha de la figura 2 se puede observar un ejemplo del segundo caso.

Si bien no es el objetivo general, muchos usuarios buscan tener cuentas populares, ya sea para que su contenido llegue a más personas, generar un posicionamiento en su marca o producto, o simplemente para sentirse importante entre los usuarios.

Cuando una cuenta desea compartir contenidos, es posible que un solo tweet no sea suficiente, por lo que deberá crear varios de forma seguida, teniendo esto en cuenta, Twitter lanzó una nueva funcionalidad llamada Threads, con la cual es posible redactar varios tweets de forma continua, los cuales luego serán asociados unos con otro para así poder ser mostrados de forma consecutiva, un ejemplo de esto se puede observar en la figura 3.

Figura 3: Ejemplo de Thread
Fuente: Twitter para Android.

2.2. Técnicas de almacenamiento de datos

Técnicas de almacenamiento de dato...

Dimensiones a estudiar

La solución al problema a desarrollar posee 3 ejes claramente distinguibles, el primero es la detección de tópico o tema tratado, el segundo es la detección de emociones, y tercero, el análisis de influencia (popularidad y viralización) de los contenidos en la red social Twitter.

2.3. Detección de Tópicos

La detección de tópicos en el análisis de lenguaje natural (NLP) hace referencia al estudio de documentos para reconocer ciertos ”Tópicos” o temas que se puedan estar tratando, es decir, es una forma de ”etiquetar” el tema en un documento según las palabras utilizadas.

Una de las técnicas utilizadas en la detección de tópicos es Latent Dirichlet Allocation (LDA), el cual es un modelo estadístico generativo que aprende los tópicos latentes presentes en una colección de documentos [Petrović et al., 2010].

Para el caso de estudio, la colección de documentos sería el conjunto de tweets pertenecientes a cada Thread. Es factible también considerar el thread de tweets estudiado como un solo documento, lo que hace posible realizar una detección de tópicos más precisa, esto debido a que es poco probable que un tweet sea lo suficientemente complejo como para detectar un tópico debido a que las limitaciones presentadas por la plataforma condicionan la calidad del lenguaje utilizado[Roberts et al., 2012], por lo que al aglomerar varios tweets en un documento más grande, el texto resultante posee más palabras asociadas al tema abstracto.

Existen muchas variantes de LDA para realizar detección de tópicos. En el documento [Petrović et al., 2010] se presenta una adaptación para detección de nuevos eventos, también conocido en ingles como First Story Detection. En el texto presentan un algoritmo que a través de un stream de tweets busca detectar la primera historia relacionada a un evento particular, el cual a través de hashing analiza la similitud entre documentos, para así generar un conjunto de cada tema detectado. Cada tema tiene un limite de documentos asociados, por lo que si se debe agregar uno nuevo, se elimina el más antiguo.

En este documento también generan una red de tweets por cada tema, al cual luego le analizan la velocidad de crecimiento, esto para estudiar el interés e impacto generado por el tema, para así diferenciar los temas más relevantes.

Otra variante presentada sobre LDA se encuentra en el texto [Lau et al., 2012],en el cual se

busca algo similar al texto mencionado anteriormente, diferenciandose en el método de obtención y análisis de documentos. En esta variante se utiliza un vocabulario dinámico, enfocado en la clusterización y análisis de co-ocurrencia más que en la frecuencia de términos. Este vocabulario se va actualizando en porciones de tiempo definidas, lo que permite mantener actualizados los tópicos.

La forma de manejar el texto en relación al stream de datos en este texto se basa en ventanas de tiempo, lo que hace que los documentos analizados se dividan en cada ventana, esto se hace para mantener un tamaño constante de documentos analizados.

2.4. Detección de Emociones

La detección de emociones es parte de un área mayor llamada Affective Computing, la cual busca que los computadores sean capaces de detectar y expresar emociones humanas [Canales y Martínez-Barco, 2014].

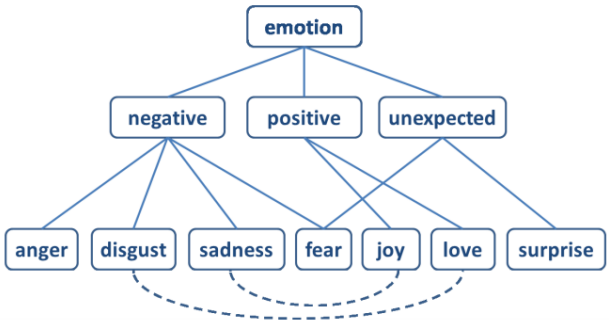


Figura 4: Relación entre las 6 emociones básicas además de amor, lineas sólidas representan herencia, mientras que las punteadas sentimientos contrarios
Fuente: Empatweet [Roberts et al., 2012].

Es posible clasificar las emociones a través de diversos modelos, poseyendo estos mayor o menor nivel de especificación en sus categorías. Dependiendo del estudio realizado podrán ser clasificadas en distintas clases. Por ejemplo, Con Sentiment Analysis es posible obtener un rango para los sentimientos entre Positivo, Negativo o Neutro, mientras que otras categorías se basan en estudios psicológicos, los cuales reconocen varias emociones básicas: Enojo, Disgusto, Miedo, Felicidad y Sorpresa [Canales y Martínez-Barco, 2014], vale decir que las emociones mencionadas pueden ser categorizadas en los setminientos también descritos, un ejemplo de esto se ve en la figura 6.

Para la detección de emociones se reconocen tres posibles métodos para realizar esto: Métodos basados en léxicos, métodos basados en máquinas de aprendizaje supervisadas y máquinas de aprendizajes no supervisadas, teniendo cada una sus ventajas y desventajas [Canales y Martínez-Barco, 2014].

Es posible construir corpus para la detección de emociones presentes en tweets a través de diversas técnicas de reconocimiento de emociones presentes en tweets a través de diversas técnicas de reconocimiento de emociones, tales como extracción y etiquetado de hashtags, detección de tópicos previamente etiquetados según emociones, máquinas de aprendizajes guiadas con tweets previamente etiquetados, entre otros [Hasan et al., 2014] [Mohammad, 2012] [Roberts et al., 2012].

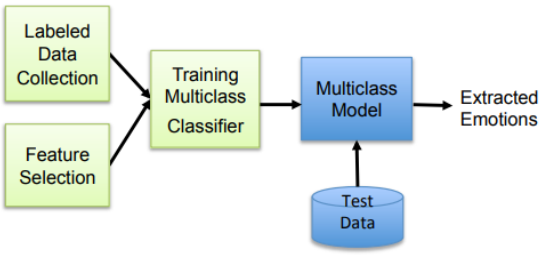


Figura 5: Ejemplo de modelo de aprendizaje supervisado para la detección de emociones en mensajes

Fuente: EmoTex [Mohammad, 2012].

Para cada tipo de modelo de detección se presentan diversos problemas [Hasan et al., 2014], ya sea el lenguaje utilizado, que le genera problemas a técnicas basadas por léxicos; la falta de etiquetas por parte del texto a analizar, que no permite entrenar máquinas de aprendizaje; e incluso la gran cantidad de temas tratados en twitter, que genera un gran número de potenciales tópicos, los que presentados en forma de vectores generaría una gran cantidad de valores cero para un tweet específico en técnicas de análisis basados en estos.

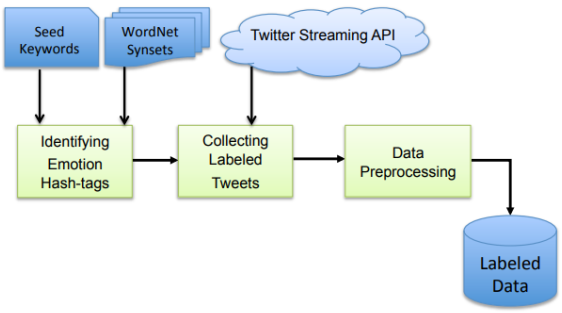


Figura 6: Ejemplo de modelo para la generación de una colección de datos etiquetados
Fuente: EmoTex [Mohammad, 2012].

2.5. Análisis de influencia

Además de analizar el contenido de los tweets en sí, también es necesario estudiar el impacto que genera. Actualmente existen varias formas de interactuar con el contenido en

Twitter, es posible Responder, compartir en forma de Retweet (RT) y poner Me gusta a los tweets(anteriormente llamado Favorito), siendo cada una de estas acciones muy distinta en la intención con la que se hacen.

Otra dimensión que hay que tomar en cuenta al momento de analizar el impacto que tiene un tweet, es la calidad e interés que genera el mismo usuario que crea el contenido, puesto que dependiendo de el interés que genere el creador, distintas serán las reacciones que generen [Cha et al., 2010].

En la literatura se reconocen 3 tipos de influencia que pueda generar un usuario [Cha et al., 2010], indegree Influence, la cual se refiere a la cantidad de seguidores que posea el usuario; Retweet Influence, la que medida según la cantidad de RTs que reciben sus publicaciones, indica la habilidad del usuario de generar contenidos con valor de ser compartidos; y Mention Influence, la cual indica el interés generado por el usuario para que otros deseen çonversarçon el.

Para cada tipo de influencia se reconocen ciertos perfiles que generan mayor tipo de interés [Cha et al., 2010]. Para Indegree existe una gran variedad de tipos de usuarios, siendo la mayoría canales de noticias, políticos, famosos y celebridades, mientras que para Retweet existe predominancia por cuentas de noticias y otros contenidos, como por ejemplo memes y videos. Finalmente las cuentas que generan mayor interés en el ámbito de las menciones, se reconoce que en su mayoría son cuentas de celebridades.

Es posible también reconocer que tan transversal es la influencia de las cuentas populares en cada una de las 3 categorías de interés. Se reconoce en la literatura que las cuentas populares en RT y menciones poseen una correlación no despreciable, tal como se puede apreciar en la figura 7, además de que las cuentas más importantes poseen un impacto elevado en diversos tópicos, mientras que cuentas menos populares poseen popularidad en temas más acotados [Cha et al., 2010].

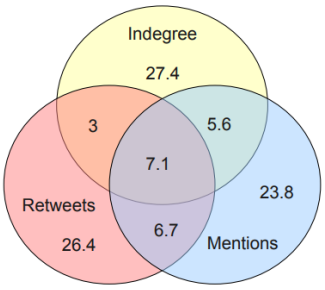


Figura 7: Diagrama de Venn de los top 100 de cada categoría de influencia, los valores se encuentran normalizados para que sumen 100 %.
Fuente: Measuring user influence in twitter. [Cha et al., 2010].

Análisis de dimensiones

Luego de completar los tweets con la información faltante, además de complementarlos con la información generada a través de los análisis desarrollados, se debe estudiar los comportamientos que esto poseen, para así detectar prácticas a repetir, con el objetivo de viralizar contenidos.

Existen diversas técnicas para el análisis de datos. Es posible realizar gráficos de los atributos de interés para visualizar y descubrir comportamientos de los datos.

También es posible desarrollar técnicas de minería de dato para descubrir relaciones ocultas entre los datos.

2.6. Métricas de popularidad

Se definirá que se considera popularidad o éxito, o alguna palabra que represente mejor el objetivo de la memoria.

2.7. Minería de datos

La minería de datos es un campo de las ciencias de la computación y estadística, la cual a través de conocimientos de inteligencia artificial, maquinas de aprendizaje, estadística y sistemas de bases de datos, se pueden descubrir patrones en grandes conjuntos de datos.

A través de la minería de datos, es posible detectar conductas como grupos (Clusters) en los datos, los cuales representan ciertas tendencias, también se pueden detectar anomalías y dependencias entre los datos.

Es posible realizar dos tipos de análisis, Descriptivo y Predictivo. El primero permite estudiar los datos en sí, evaluar la distribución y tendencias, mientras que el segundo es utilizado para modelar fenómenos, entrenar máquinas de aprendizaje automático, entre otros. El análisis predictivo es utilizado principalmente para hacer prediciones de comportamientos futuros o no conocidos.

Existen diversas técnicas de minería de datos, las cuales pueden realizar análisis descriptivos o predictivos, siendo las más importantes las siguientes:

- Redes Neuronales.
- Arboles de decisión
- Clustering

- Reglas de Asociación
- Regresión Lineal
- Modelos Estadísticos

2.8. Metodologías existentes

Metodologías existentes para poder realizar el trabajo deseado.

CAPÍTULO 3
PROPUESTA DE SOLUCIÓN

3.1. Metodología de trabajo

Definir metodología a utilizar. Pasos, etapas y repeticiones.

3.1.1. Almacenamiento de datos

Luego de obtener obtener los datos a través de las APIs, se posee un tweet con los siguientes atributos (id, thread, timestamp, contenido, rt, likes, replies, id usuario,usuario es verificado?,cantidad de seguidores usuario, cantidad de seguidos usuario, cantidad de tweets usuario).

Además, luego del procesamiento del contenido de los tweets a través de análisis de tópico y sentimiento, se obtiene nueva información del tweet (id, emociones detectadas, tópicos detectados).

3.1.2. Detección de tópicos

Técnica a usar.

3.1.3. Reconocimiento de emociones

Técnica a usar.

3.1.4. Influencia del usuario

Metodología a usar.

3.1.5. Popularidad

Fórmula para cuantificar.

3.1.6. Análisis de datos

Técnica de minería de datos.

3.2. Análisis de tópico

En este punto se desarrollará el estudio de tópicos según Hilos, se estudiarán los tópicos que más aparecen, el largo del hilo por tópico, etc.

Relación entre tópico y popularidad

Se estudiará la distribución de los tópicos y la popularidad

3.3. Análisis de Emociones

En este punto se analizarán las emociones por hilo, la distribución de la emociones según el largo, etc.

Relación entre emociones y popularidad

Se estudiará la distribución de la popularidad de las emociones.

3.4. Análisis de Influencia

Acá se estudiará la relación directa entre el perfil del usuario y la popularidad del Hilo.

Relación entre la influencia y la popularidad

3.5. Relación entre dimensiones

En esta sección se estudiarán las relaciones de todas las combinaciones de los 3 análisis, luego se estudiará la relación entre las dos dimensiones más concluyentes

3.5.1. Relación entre tópico y emoción

3.5.2. Relación entre tópico e influencia

3.5.3. Relación entre influencia y emoción

3.6. Análisis de relaciones obtenidas

Hacer comentarios de los resultados obtenidos

3.7. Selección y construcción de recomendaciones

Indicar cuales son las relaciones que mas impactan, indicar que funciona y que no.

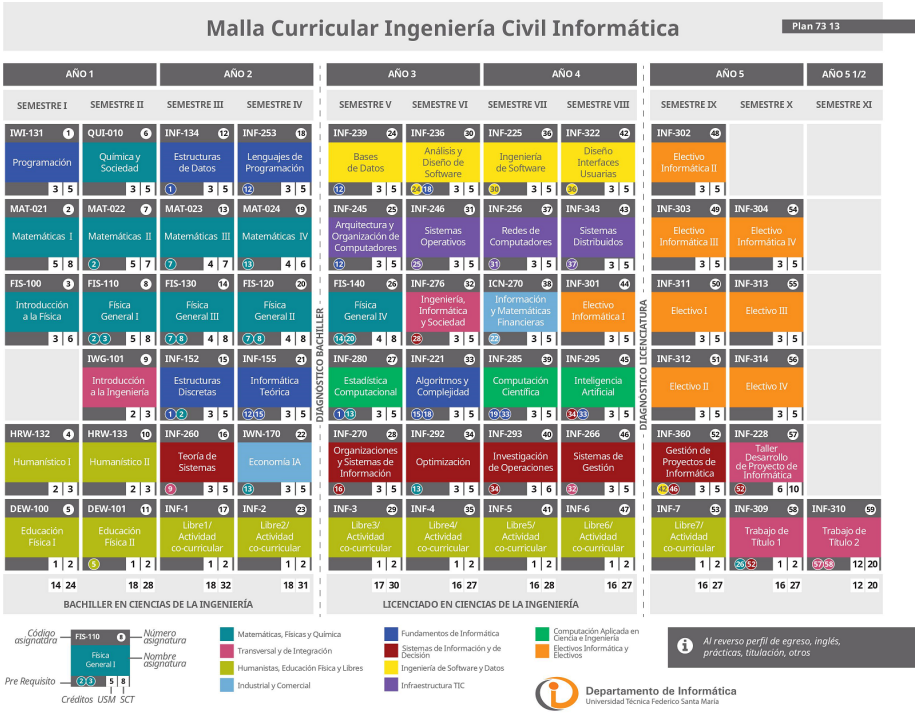


Figura 8: Malla Curricular Ingeniería Civil Informática.
Fuente: Departamento de Informática.

CAPÍTULO 4
VALIDACIÓN DE LA SOLUCIÓN

La validación se realizará analizando otro set de datos y se evaluará si las observaciones realizadas en el punto anterior se cumplen

4.1. Datos

Datasets a utilizar

4.2. Pruebas

Análisis de los datos

Tabla 1: Coloquios del Ciclo de Charlas Informática.
Fuente: Elaboración Propia.

Título Coloquio	Presentador, País
"Sensible, invisible, sometimes tolerant, heterogeneous, decentralized and interoperable... and we still need to assure its quality..."'	Guilherme Horta Travassos, Brasil.
"Dispersed Multiphase Flow Modeling: From Environmental to Industrial Applications"'	Orlando Ayala, EE.UU.
"Líneas de Producto Software Dinámicas para Sistemas atentos el Contexto"'	Rafael Capilla, España.
...	...

CAPÍTULO 5
CONCLUSIONES

Las Conclusiones son, según algunos especialistas, el aspecto principal de una memoria, ya que reflejan el aprendizaje final del autor del documento. En ellas se tiende a considerar los alcances y limitaciones de la propuesta de solución, establecer de forma simple y directa los resultados, discutir respecto a la validez de los objetivos formulados, identificar las principales contribuciones y aplicaciones del trabajo realizado, así como su impacto o aporte a la organización o a los actores involucrados. Otro aspecto que tiende a incluirse son recomendaciones para quienes se sientan motivados por el tema y deseen profundizarlo, o lineamientos de una futura ampliación del trabajo.

Todo esto debe sintetizarse en al menos 5 páginas.

ANEXOS

En los Anexos se incluye todo aquel material complementario que no es parte del contenido de los capítulos de la memoria, pero que permiten a un lector contar con un contenido adjunto relacionado con el tema.

REFERENCIAS BIBLIOGRÁFICAS

[Canales y Martínez-Barco, 2014] Canales, L. y Martínez-Barco, P. (2014). Emotion detection from text: A survey. En Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC), pp. 37–43.

[Cha et al., 2010] Cha, Meeyoung and Haddadi, Hamed and Benevenuto, Fabricio and Gummadi, P Krishna and others (2010). Measuring user influence in twitter: The million follower fallacy. Icwsm, 10(10-17):30.

[Hasan et al., 2014] Hasan, M., Rundensteiner, E., y Agu, E. (2014). Emotex: Detecting emotions in twitter messages.

[Lau et al., 2012] Lau, J. H., Collier, N., y Baldwin, T. (2012). On-line trend analysis with topic models:\# twitter trends detection topic model online. Proceedings of COLING 2012, pp. 1519–1534.

[Mohammad, 2012] Mohammad, S. M. (2012). # emotional tweets. En Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, pp. 246–255. Association for Computational Linguistics.

[Petrović et al., 2010] Petrović, S., Osborne, M., y Lavrenko, V. (2010). Streaming first story detection with application to twitter. En Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics, pp. 181–189. Association for Computational Linguistics.

[Roberts et al., 2012] Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., y Harabagiu, S. M. (2012). Empatweet: Annotating and detecting emotions on twitter. En LREC, volumen 12, pp. 3806–3813. Citeseer.