



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA  
La Universidad Católica de Loja

ÁREA TÉCNICA

TITULACIÓN DE INGENIERO EN SISTEMAS INFORMÁTICOS Y  
COMPUTACIÓN

Estudio de los cambios en la descripción de recursos publicados en redes sociales.  
Aplicación en twitter

TRABAJO DE FIN DE TITULACION

**AUTORA:**

Agila Navarro, Gabriela María.

**DIRECTOR:**

López Vargas, Jorge Afranio, Ing.

LOJA –ECUADOR

2014

## **APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN**

Ingeniero.

Jorge Afranio López Vargas.

### **DOCENTE DE LA TITULACIÓN**

De mi consideración:

El presente trabajo de fin de titulación: Estudio de los cambios en la descripción de recursos publicados en redes sociales. Aplicación Twitter, realizado por Agila Navarro Gabriela María, ha sido orientado y revisado durante su ejecución, por cuanto se aprueba la presentación del mismo.

Loja, abril de 2014

f).....

## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“Yo Agila Navarro Gabriela María declaro ser autor (a) del presente trabajo de fin de titulación: Estudio de los cambios en la descripción de recursos publicados en redes sociales. Aplicación Twitter, de la titulación de Ingeniera en Sistemas Informáticos y Computación, siendo Jorge Afranio López Vargas director del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través, o con el apoyo financiero, académico o institucional (operativo) de la Universidad”

f. ....

Autor: Agila Navarro Gabriela María

Cédula: 1104639024

## DEDICATORIA

Con todo cariño dedico este trabajo a personas importantes en mi vida, que siempre estuvieron ahí para brindarme toda su ayuda y apoyo incondicional, a quienes son mi motor y motivo de inspiración, a quienes me enseñaron que la inteligencia es la fuente de un hombre próspero y que estudiar es un valor incalculable de la vida, y que de cierta forma han estado en los buenos y malos momentos junto a mí.

A mis padres Ecuador Bolívar y María Elsa.  
A mis hermanos Cinthya Melania y Alexander Ronaldo.

## **AGRADECIMIENTO**

A Dios y la Virgen del Cisne por regalarme la vida, sabiduría y la oportunidad de tomar este camino lleno de expectativas y deseos de triunfar.

A mis padres que son mi ejemplo de superación y entrega, porque en gran parte gracias a ustedes, hoy puedo ver alcanzada mi meta, ya que siempre estuvieron impulsándome en los momentos difíciles, enseñándome a encarar las adversidades sin perder nunca la dignidad ni desfallecer en el intento.

A mis hermanos que son mi motor y motivo de inspiración para cumplir mis objetivos, quienes me han acompañado en cada paso que doy, dándome fortaleza y ánimos para continuar.

A esa persona especial por darme las fuerzas para seguir adelante y no desmayar en los problemas que se presentaban, por su apoyo incondicional, su paciencia y comprensión.

A mi familia gracias por haber fomentado en mí el deseo de superación y el anhelo de triunfo en la vida, al brindarme sus consejos.

A mis amigos con quienes tengo el gusto de compartir momentos amenos, quienes me incentivaron para seguir adelante.

A mi Director de Tesis por guiarme en cada paso de este proyecto, aportándome sus conocimientos y sugerencias, gracias por su asesoría, por su contribución y predisposición en el desarrollo de la presente investigación.

A todos mis profesores por impartirme sus enseñanzas y experiencias, por prepararnos para un futuro competitivo no solo como los mejores profesionales sino también como mejores personas.

A la Universidad Técnica Particular de Loja (UTPL), y a todo el laboratorio de Tecnologías Avanzadas en la Web y SBC por la oportunidad para desarrollar la investigación.

## ÍNDICE DE CONTENIDOS

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN .....	ii
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS.....	iii
DEDICATORIA .....	iv
AGRADECIMIENTO .....	v
ÍNDICE DE CONTENIDOS .....	vi
ÍNDICE DE TABLAS .....	ix
ÍNDICE DE FIGURAS.....	xi
RESUMEN.....	1
ABSTRACT .....	2
INTRODUCCIÒN.....	3
ESTADO DEL ARTE.....	5
1.1. Difusión de la información.....	6
1.2. Twitter.....	7
1.2.1. Origen de twitter .....	8
1.2.2. Twitter como herramienta de difusión .....	9
1.2.3. Modelo comunicativo de twitter.....	13
1.2.4. Metadatos de un tweet .....	14
1.2.5. Tipos de metadatos .....	20
1.2.6. Datos dentro del contenido de un tweet.....	20
1.2.7. Estructura del tweet.....	23
1.2.8. Indicadores twitter .....	25
1.3. Análisis de redes sociales.....	26
1.4. Problemática.....	27
1.5. Solución.....	27
1.6. Trabajos relacionados con la difusión de información .....	28

1.7. Trabajos relacionados con twitter. ....	29
DISEÑO E IMPLEMENTACIÓN DEL ALGORITMO.....	31
2.1. Representación de algoritmos .....	32
2.1.1. Diagrama de flujo .....	32
2.1.2. Diagrama Nassi-Shneiderman.....	32
2.1.3. Pseudocódigo.....	33
2.2. Comparativa de las notaciones algorítmicas.....	33
2.3. Algoritmo .....	35
2.3.1. Entorno de desarrollo NetBeans.....	35
2.3.2. Servidor de base de datos .....	35
2.3.3. Descripción del proceso del algoritmo .....	36
2.3.3.1. Identificación de datos .....	38
2.3.3.2. Obtención de urls.....	38
2.3.3.3. Recuperar tweets.....	40
2.3.3.4. Recuperar etiquetas, menciones, urls, texto.....	41
2.3.4. Determinar cambios.....	53
2.3.5. Almacenar resultados.....	56
2.3.6. Presentación de resultados .....	57
ESTUDIO REAL DE LOS CAMBIOS .....	58
3.1. Cambios de las urls que compartieron el mayor número de tweets. ....	61
3.2. Cambios de las urls que compartieron el menor número de tweets. ....	64
3.3. Urls con mayor y menor número de cambios en menciones.....	66
3.4. Urls con mayor y menor número de cambios en etiquetas.....	69
3.5. Urls con mayor y menor número de cambios en urls .....	72
3.6. Urls con mayor y menor número de cambios en el texto.....	75
3.7. Urls con mayor y menor promedio en cambios .....	78
3.8. Promedio de los resultados en todas las métricas .....	82

CONCLUSIONES.....	83
RECOMENDACIONES.....	85
TRABAJOS FUTUROS.....	86
BIBLIOGRAFÍA.....	87
ANEXOS.....	93



## ÍNDICE DE TABLAS

Tabla 1: Notaciones algorítmicas para la representación de algoritmos .....	33
Tabla 2: Datos para el estudio de los cambios .....	38
Tabla 3: Etiquetas concatenadas por el TWETT_ID de la URL_ID=1 .....	42
Tabla 4: Menciones concatenadas por el TWETT_ID de la URL_ID=1 .....	44
Tabla 5: Urls cortas y Urls extendidas .....	45
Tabla 6: Urls extendidas de los tweets que compartieron la URL_ID=1 .....	46
Tabla 7: Extraer menciones .....	48
Tabla 8: Extraer etiquetas .....	49
Tabla 9: Extraer urls .....	50
Tabla 10: Signos a reemplazar .....	51
Tabla 11: Extraer signos .....	51
Tabla 12: Texto de los tweets que compartieron la URL_ID=1 .....	52
Tabla 13: ArrayList etiquetas.....	53
Tabla 14: Número de repeticiones de etiquetas .....	54
Tabla 15: ArrayList de etiquetas.....	54
Tabla 16: Número de cambios .....	55
Tabla 17: Total de cambios en etiquetas, menciones, urls y texto de la URL_ID=1 .....	55
Tabla 18: Promedio de los cambios .....	55
Tabla 19: Resultados de los cambios de la URL_ID=1 .....	56
Tabla 20: Cálculo del rango .....	60
Tabla 21: Cálculo de medio rango .....	60
Tabla 22: Cambios de las 20 urls que se compartieron en más tweets .....	61
Tabla 23: Urls que compartieron el mayor número de tweets .....	61
Tabla 24: Cambios de la URL_ID =3365 .....	62
Tabla 25: Porcentaje de cambios de la URL_ID=3365.....	63
Tabla 26: Cambios de las 20 urls que compartieron el menor número de tweets.....	64
Tabla 27: Urls que compartieron el menor número de tweets .....	64
Tabla 28: Cambios de la URL_ID =4 .....	65
Tabla 29: Urls con mayor y menor número de cambios en menciones.....	66
Tabla 30: Url con mayor número de cambios en mención .....	68
Tabla 31: Url con menor número de cambios en mención .....	68
Tabla 32: Urls con mayor y menor número de cambios en etiquetas.....	69
Tabla 33: Url con mayor número de cambios en etiquetas .....	71
Tabla 34: Url con menor número de cambios en etiquetas .....	71
Tabla 35: Urls con mayor y menor número de cambios en urls .....	72
Tabla 36: Url con mayor número de cambios en urls .....	74

Tabla 37: Url con menor número de cambios en urls .....	74
Tabla 38: Urls con mayor y menor número de cambios en el texto.....	75
Tabla 39: Url con mayor número de cambios en el texto .....	77
Tabla 40: Url con menor número de cambios en el texto .....	77
Tabla 41: Urls con mayor y menor número de promedio en cambios. ....	78
Tabla 42: Url con mayor promedio en cambios .....	80
Tabla 43: Url con menor promedio en cambios .....	80
Tabla 44: 10 urls con mayor cambio en el promedio de los tweets. ....	81
Tabla 45: 10 urls con menor cambio en el promedio de los tweets.....	81
Tabla 46: Promedio de los resultados de todas las métricas .....	82

## ÍNDICE DE FIGURAS

Figura 1: Contribuyentes de un tweet.....	15
Figura 2: Coordenadas de un tweet.....	15
Figura 3: Fecha de creación de un tweet.....	15
Figura 4: Lenguaje de un tweet.....	16
Figura 5: Tweet favorito.....	16
Figura 6: Número de favoritos de un tweet.....	16
Figura 7: Tweet retwitteado.....	16
Figura 8: Número de retweets.....	17
Figura 9: Lugar de un tweet.....	17
Figura 10: Entidades de un tweet.....	17
Figura 11: Hashtags de un tweet.....	18
Figura 12: Urls de un tweet.....	18
Figura 13: Menciones de un tweet.....	18
Figura 14: Elementos multimedia de un tweet.....	18
Figura 15: Texto de un tweet.....	19
Figura 16: Usuario de twitter.....	19
Figura 17: Etiquetas de un tweet.....	21
Figura 18: Mención de un tweet.....	21
Figura 19: Texto de un tweet.....	22
Figura 20: Link de un tweet.....	22
Figura 21: Url acortada de un tweet.....	23
Figura 22: Estructura simple de un tweet.....	23
Figura 23: Estructura compuesta de un tweet.....	24
Figura 24: Diagrama de flujo del algoritmo.....	34
Figura 25: Proceso general para determinar los cambios.....	36
Figura 26: Tareas del las fases del proceso de cambios.....	37
Figura 27: Urls extendidas a ser analizadas.....	38
Figura 28: Tweets que comparten una misma URL.....	39
Figura 29: Tabla tweet.....	40
Figura 30: Tweets que comparten la URL_ID=1.....	40
Figura 31: Tweets que comparten la URL_ID=3.....	41
Figura 32: Tweets que comparten la URL_ID=4.....	41
Figura 33: Tabla etiquetas.....	41
Figura 34: Etiquetas de los tweets.....	42
Figura 35: Tabla menciones.....	43
Figura 36: Menciones de los tweets.....	43

Figura 37: Tabla urls .....	44
Figura 38: Índices de las menciones.....	48
Figura 39: Índices de etiquetas .....	49
Figura 40: Índices de urls .....	50
Figura 41: Tabla gp_resultados.....	56
Figura 42: Interfaz para presentar los resultados .....	57
Figura 43: Resultados de los cambios .....	59
Figura 44: Fórmula de rango.....	60
Figura 45: Fórmula de medio rango .....	60
Figura 46: Fórmula de la media aritmética.....	60
Figura 47: Fórmula de la desviación promedio.....	60
Figura 48: Cambios de las urls que han compartido el mayor número de tweets.....	62
Figura 49: Porcentaje de cambios de la URL=3365 .....	63
Figura 50: Cambios de las urls que han compartido el menor número de tweets.....	65
Figura 51: Urls con el mayor número de cambios en menciones.....	67
Figura 52: Urls con el menor número de cambios en menciones.....	67
Figura 53: Porcentaje de cambios de la URL=180 .....	68
Figura 54: Urls con el mayor número de cambios en etiquetas.....	70
Figura 55: Urls con el menor número de cambios en etiquetas .....	70
Figura 56: Porcentaje de cambios de la URL=17236 .....	71
Figura 57: Urls con el mayor número de cambios en urls .....	73
Figura 58: Urls con el menor número de cambios en urls .....	73
Figura 59: Porcentaje de la URL_ID=3 .....	74
Figura 60: Urls con el mayor número de cambios en el texto.....	76
Figura 61: Urls con el menor número de cambios en el texto .....	76
Figura 62: Porcentaje de cambios de la URL=2416 .....	77
Figura 63: Urls con el mayor número de promedio de los cambios.....	79
Figura 64: Urls con el menor número de promedio de los cambios .....	79
Figura 65: Promedio de cambios de la URL=36882 .....	80
Figura 66: Porcentaje del promedio de cambios en todas las métricas .....	82
Figura 67: Diagrama de clases .....	93
Figura 68: Modelo relacional de la base de datos .....	94

## **RESUMEN**

En este trabajo se presenta un estudio sobre los cambios en la descripción de recursos publicados como links en la red social Twitter, los mismos que fueron medidos a través de etiquetas, menciones, urls y texto. Con esto se pretende diseñar e implementar un algoritmo que determine los cambios que tiene un recurso (expresado como una url y compartido en Twitter), para realizar un estudio real de la difusión de información en esta red, de manera que los resultados obtenidos permitan conocer el número de cambios que experimenta un recurso en actividades propias de Twitter como re-tweet y otras co-ocurrencia de URL.

Palabras claves: Etiquetas, Menciones, Urls, Texto, SNA, Difusión de la Información, Twitter.

## **ABSTRACT**

In this work it is presented a study about the changes in the description of published resources as links in the social network Twitter, the same have been measured through tags, mentions, urls, and text. With this, it is intended to design and implement an algorithm that determines the changes that a has resource (expressed as an url and shared on Twitter) to carry out a real study of the diffusion of information in this network, so that, the obtained results allow to know the number of changes that a resource experiment in proper activities of Twitter such as re-tweet and other co-occurrence URL.

Keywords: Tags, mentions, Urls, Text, SNA, Information Dissemination, Twitter.

## INTRODUCCIÓN

Los medios de comunicación revolucionaron la vida de las personas a medida que fueron apareciendo en su contexto de emergencia, esto ocurrió en nuestro país y en el mundo entero. Internet no ha sido la excepción, desde su surgimiento cambió rotundamente la forma de comunicación que se manejaba hasta entonces, fue evolucionando a pasos agigantados hasta convertirse en lo que hoy conocemos la web 2.0 que representa la transición de las páginas tradicionales a las aplicaciones web orientadas a los usuarios, estos sitios web permiten compartir información actuando como puntos de encuentro globales.

Las redes sociales en internet son comunidades virtuales donde sus usuarios interactúan con personas de todo el mundo con quienes encuentran gustos o intereses en común. Funcionan como una plataforma de comunicaciones que permite conectar gente que se conoce o que desea conocerse, y que les permite centralizar recursos, como fotos y vídeos, en un lugar fácil de acceder y administrado por los mismos usuarios. Entonces una red social es una estructura social conformada por grupos de personas las cuales están conectadas por uno o varios tipos de relaciones formando vínculos entre sus usuarios.

Twitter es una red que permite enviar mensajes de texto plano de corta longitud, con un máximo de 140 caracteres, llamados tweets, se promociona como una de las mejores maneras de compartir y descubrir qué está pasando a nuestro alrededor en este preciso momento, cabe destacar que no sólo proporciona información relacionada sobre personalidades, noticias, recursos, entre otros, sino que también permite seguir estas interacciones, es por esto que se trata de comprender que usos se le da hoy a esta red, sobre todo en la comunicación, la transmisión, análisis y manipulación de información.

El objetivo principal de este proyecto es realizar un estudio de los cambios en la descripción de recursos que contengan links publicados en la red social, específicamente twitter en donde se determinará el cambio del “tweet” que será medido mediante varias métricas definidas como: “etiquetas”, “menciones”, “urls” y “texto”. Y para dar cumplimiento al mismo se propone diseñar e implementar un algoritmo que determine los cambios que experimenta un recurso, para finalmente poder realizar un estudio real de la difusión de información en twitter.

El presente trabajo consta de 3 capítulos:

- 1) Primer capítulo, abarca el Estado del Arte que brinda un enfoque general sobre: análisis de redes sociales, difusión de la información, descripción de twitter, modelo comunicativo de twitter, datos dentro del contenido de un tweet, estructura de un tweet.
- 2) Segundo capítulo, trata sobre el Diseño e Implementación del Algoritmo, en este se plantea el algoritmo, la descripción del proceso que se debe llevar a cabo para determinar los cambios de un recurso.
- 3) Tercer capítulo presenta un Estudio Real de los Cambios, el mismo que detalla un análisis de los resultados que se obtuvieron al ejecutar el Algoritmo, se podrá visualizar gráficas estadísticas de los cambios encontrados.

Finalmente se expone las conclusiones y recomendaciones recogidas a partir del análisis de los datos de los tweets, del diseño y la implementación del algoritmo, y de los cambios observados al ejecutarlo en un contexto definido.

Con los resultados obtenidos a partir del desarrollo de este proyecto, se tiene la expectativa de ampliar el campo de investigación sobre la información que se genera a partir de no sólo las redes sociales en la web, sino también de muchos otros campos como pueden ser:

- Ámbito educacional
- Recursos informativos
- Recursos bibliográficos
- Datos médicos
- Entre otros

Esto con el fin de enriquecer los procesos de investigación llevados a cabo en la universidad y fortalecer el desarrollo del conocimiento propiciado por la misma.



## **CAPÍTULO 1**

### **ESTADO DEL ARTE**

## 1.1. Difusión de la información

La información es poder sólo cuando la información es compartida y produce conocimiento colectivo susceptible de mejorar cualquier aspecto de la sociedad. En este sentido las redes sociales juegan un papel revolucionario dónde perderse en el anonimato es cada vez más fácil. (Orta, Pardo, & Salas, 2012)

Se denomina difusión de la información o broadcasting al proceso mediante el cual la información que posee un nodo fuente es distribuida al resto de los nodos de la red, en el modelo clásico uno a uno, en cada etapa un nodo puede enviar el mensaje a lo sumo a uno de sus vecinos y tarda en realizar este proceso una unidad de tiempo. (Amblás, Bendito, Carmona, Encinas, Joan, & Mitjana, 2003)

Potencialmente, la difusión de información en las redes sociales puede ser muy eficiente y la información puede llegar a un gran número de personas en un período muy corto de tiempo. (Holmberg, 2012). Sin embargo, poco se sabe acerca de los patrones de difusión de información en los medios sociales y cómo el comportamiento de información en línea puede haber cambiado.

Twitter tiene una amplia colaboración para la difusión de información, mediante sus servicios, la difusión del hashtag y el retweet como tema está gobernada por la influencia de aquellos que lo utilizan. (Romero, Meeder, & Kleinberg, 2011) Esto es una gran ventaja porque ayuda a organizar la información para que pueda ser difundida y buscada de mejor manera.

La influencia se traduce en la capacidad de hacer llegar el mensaje al mayor número de individuos posibles, donde la información en twitter fluye de manera viral a partir de la creación y generación de un mensaje y su posterior redifusión por la red de influencia de la persona que lo emite, y de las redes de los que lo reciben y deciden reemitirlo, y así sucesivamente hasta que el mensaje muere (Orta, Pardo, & Salas, 2012). También se puede definir "influencia" como la probabilidad asociada a un nodo de transmitir o impedir la transmisión de nuevas ideas o pautas de comportamiento en la red.

Los expertos coinciden que la aparición de Internet y los navegadores son un hito en el ámbito de la difusión de información. Por ejemplo, la web 2.0 ha contribuido en el desarrollo de

sistemas descentralizados de distribución de información, como las páginas wiki (Wikipedia<sup>1</sup>), las bitácoras o microblogs (Twitter<sup>2</sup>), los sistemas de alertas (RSS) o los sistemas de votación (Menéame<sup>3</sup>) han permitido la construcción de comunidades virtuales que aprovechan la inteligencia colectiva para generar conocimiento y difundirlo. (De la Rosa Troyano, 2012).

En otras palabras, los medios sociales hacen referencia a servicios de internet que son importantes en la vida cotidiana, donde las personas pueden interactuar unos con otros, compartir y crear contenidos para la web, pero la manera en que se utiliza la misma está cambiando. (Holmberg, 2012). Y en la actualidad el énfasis se ha desplazado de la búsqueda de información hacia el intercambio de la misma (Kwak, Lee, Park, & Sue, 2010), ya que la web que solía ser un lugar donde se realizan búsquedas de información, hoy también es un lugar para compartir información e interactuar con otras personas, convirtiéndose así en una gran fuente de datos.

De esta manera el desarrollo de la Web 2.0 y la aparición de redes sociales han supuesto una revolución en el ámbito de la comunicación que ha cambiado los modelos tradicionales y ha dotado de nuevos roles a los emisores, que no se limitan a recibir información sino también a crearla y difundirla. (López Abellán, 2012)

Debido a la rápida aceptación de estas nuevas herramientas de difusión y comunicación cada día millones de usuarios se conectan a través de redes sociales, generando un rico tesoro de datos que permiten estudiar los mecanismos que subyacen a las interacciones humanas, los accesos directos basados en el tráfico son otro factor clave en la interpretación de la evolución de la red. Los usuarios que son populares, activos e influyentes tienden a crear accesos directos por el tráfico (Weng, y otros, 2013), haciendo que el proceso de difusión de la información sea más eficiente en la red.

## **1.2. Twitter**

Es una red de información en tiempo real, siendo el servicio de microblogging más popular y extendido en la Red, en Twitter existen pequeñas explosiones de información llamadas Tweets (mensajes) que se publican de no más de 140 caracteres de longitud, que te conectan con las últimas historias, ideas, opiniones y noticias sobre lo que encuentras interesante como: cine,

---

<sup>1</sup> <http://www.wikipedia.org/>

<sup>2</sup> <https://twitter.com/>

<sup>3</sup> <http://www.meneame.net/>

gobierno y política, gastronomía, organizaciones, entretenimiento, fútbol, música, cuentas oficiales, elecciones, innovación, turismo, deportes, medios y periodistas, literatura, moda y belleza. Para esto se debe localizar las cuentas que te resulten más atractivas y podrás seguir sus conversaciones, generando en un punto de encuentro global. (Twitter, 2013).

Se compone de followers y following, que no tienen por qué coincidir, un usuario sigue a los usuarios que considera de su interés y es seguido por usuarios que le consideran a él por su parte también de interés. Intrínsecamente estas redes se verán conformadas por usuarios de intereses similares, bien sean políticos, económicos, de ocio, entre otros, lo que se denomina temas o grupos semánticos primarios. (Orta, Pardo, & Salas, 2012)

Por defecto, los mensajes son públicos, los usuarios pueden leer los tweets sin aprobaciones, construyendo de ese modo una red dirigida entre ellos (Kawamoto, 2012). Asimismo se puede propagar tweets hacia sus seguidores gracias a una función característica llamada retweets (Yang, y otros, 2010), mediante el cual se da la difusión de la información.

El éxito de Twitter se basa en la simplicidad de su idea matriz, twitter es una de las redes sociales favoritas de internet, es muy potente para la difusión y multiplicación rápida de titulares y enlaces a otras fuentes. Su potencial ha crecido de tal forma que actualmente lo utilizan millones de usuarios en todo el mundo para promocionarse o para utilizarlo como canal alternativo de información (Domínguez, 2012), esta red alberga cada día cientos de millones de mensajes, y vale destacar que el uso del internet como fuente de información ha desplazado casi en su totalidad a los medios de comunicación.

Twitter ha creado lo que se considera una red de informadores sociales, aquí se conjuntan dos cosas: el uso de Twitter a través de dispositivos con internet (Chang, 2010) y las noticias, como hechos ocurridos en un tiempo y lugar determinado. La red social Twitter sin duda ya forma parte de los medios de comunicación masivos, su inmediatez con la que se dan a conocer cifras o información como nombramientos, obras, campañas de difusión, actividades culturales, deportivas y muchos otros temas, es notoria para quienes hacen uso de esta red social. (Meunier Rosas, 2012)

### **1.2.1. Origen de twitter**

Twitter nació en el año 2006, aunque popular desde marzo de 2007, esta red fue creada por Jack Dorsey, Biz Stone y Evan Williams jóvenes emprendedores que trabajaban para la

compañía de Podcasts Odeo, Inc., de San Francisco, Estados Unidos. La compañía Odeo, recién había hecho una gran contribución al código de Rails 1.0 y había Odeo Studio, sin embargo estaban enfrentando una gran competencia por parte de Apple y otros. Por lo que se vieron obligados a reinventarse y en el marco de esas reuniones, Jack Dorsey propuso una idea en la que se podrían usar SMS para mantener informado a un grupo de gente sobre qué estaba uno haciendo, por ende Dorsey<sup>4</sup> es el autor de uno de los primeros borradores de Twitter que comenzó siendo un desarrollo interno de la compañía Odeo para agilizar las comunicaciones entre sus empleados. (Ramos Peinado , 2012)

Una vez iniciado el proyecto probaron varios nombres, el nombre original durante un tiempo fue “Status” (Stat.us), pasando por twitch (tic) a causa del tipo de vibraciones de los móviles, pero se quedaron con “Twitter”, que en palabras de Dorsey era perfecta, y la definición era “una corta ráfaga de información intrascendente”, el “pio de un pájaro” (T2O, 2010). Twitter está escrito en Ruby on Rails, dispone de una API abierta que permite integrar Twitter en aplicaciones web, de escritorio o móviles y entre sus principales funcionalidades están el requerimiento, retransmisión o intercambio de opiniones de eventos en directo. Recientemente el equipo de ingenieros de Twitter dieron a conocer el desplazamiento de Ruby On Rails (RoR) en su arquitectura por la tecnología Java, haciendo uso de Blender a través de Netty, argumentando mejoras de escalabilidad y rendimiento a largo y corto plazo. (Barcia Palacios, 2011)

### **1.2.2. Twitter como herramienta de difusión**

La mayoría de las personas no quieren quedarse al margen de las nuevas posibilidades comunicativas que brinda la Red, y de entre todas estas posibilidades, sin duda las que más impacto están cobrando en los últimos años son las que propician las redes sociales a la hora de interactuar con personas, creando comunidades y estableciendo un modelo de comunicación bidireccional.

De todas estas redes, Twitter es la que actualmente más relevancia ha adquirido. En palabras de Piscitelli, se ha convertido “*en uno de los mecanismos de comunicación más poderosos de la historia*” (Piscitelli, 2011). Y según Orihuela, “*ha cambiado la red y ha completado el giro social que iniciaron los blogs a finales de los años noventa*”. (Orihuela J. L., 2013)

---

<sup>4</sup> Origen de Twitter Tomado de <http://www.domisfera.com/el-origen-de-twitter/>

La brevedad es, por tanto, junto con la rapidez en la emisión y recepción de los tweets, una de las cualidades principales de esta red, a la que muchos auguraron poco éxito en sus inicios y que, muy al contrario, está alcanzando popularidad e impacto informativo hasta ahora desconocidas en otros medios y soportes de comunicación.

A continuación se indica diez razones para usar Twitter (Rodríguez & Ureña, 2012), como una herramienta en la comunicación y difusión de información.

### **Twitter aporta imagen de modernidad.**

La mayoría de las personas se suman a avances tecnológicos por el hecho de no quedarse atrás con respecto a lo que se considera innovador en la sociedad (Rodríguez & Ureña, 2012). Los usuarios que quieran influir, además de mantener una relación directa y personal, tendrán que estar, lo quieran o no, en las redes sociales". (Alcat, 2011)

### **Permite la comunicación.**

La comunicación tradicional consistía en elaborar mensajes que los difundían a través de los medios, como la prensa, la radio o la televisión. Este modelo, se sustentaba en un tipo de comunicación preferentemente unidireccional, donde solo habla la persona y la audiencia escuchaba o leía los mensajes vertidos y actuaba en consecuencia, pero sin tener apenas oportunidad de participar. Twitter, ha venido a romper estas barreras favoreciendo la comunicación, buscando nuevas maneras de tener una interacción productiva con los usuarios. Puesto que las personas pueden dirigirse directamente a un gran número de usuarios sin tener que contar con intermediadores y en donde las dos partes participantes tienen las mismas posibilidades de interactuar.

### **Los usuarios son "líderes de opinión" en sus entornos.**

Los "tuiteros" pueden ser considerados "líderes de opinión" o "influentials", puesto que son gente informada, involucrada política y socialmente, gente que se organiza y que es activa a la hora de generar debate, gente en definitiva que disfruta compartiendo información, y que trata de ejercer influencia sobre sus círculos y entornos más cercanos, como familiares, amigos, compañeros de trabajo, entre otros. Y que, por encima de todo, concibe la participación como algo continuo en el tiempo.

### **Permite la difusión y genera comunidad.**

Es útil para diseminar y difundir mensajes clave generando corrientes de opinión, esta labor de difusión se amplía de forma exponencial, llegando hasta el último usuario, que puede recibir esos mensajes, utilizarlos en sus entornos e, incluso, hacer labor de proselitismo con sus propios seguidores a través del retuiteo, generando un efecto cascada que puede lograr una extensa propagación en muy poco tiempo (Fabrega & Paredes, 2013), con una rapidez que sería impensable a través de cualquier otro medio.

Por tanto, Twitter sirve para mantener informados a las personas y para que éstas, a su vez, actúen como canal de transmisión de los mensajes (Zhao & Rosson, 2009) , en donde *“cada usuario tiene una capacidad de propagación proporcional a su número de seguidores (followers), pero el mensaje puede ser retransmitido (RT) por los followers de sus followers sin ninguna limitación, lo que hace de Twitter una de las mayores fuentes de propagación de la información en tiempo real”*. (Congosto, Fernández, & Moro Egido, 2011)

Por otro lado, estos nexos que se crean gracias a esta red pueden conformar también comunidades entre los propios seguidores, generando una mayor interconexión y relación entre ellos. (Gladwell, 2010)

### **Es un medio con gran acogida**

Twitter permite saltar las barreras que imponían los productos informativos tradicionales, no hace falta comprar un periódico o disponer de un receptor de radio o de televisión para conocer las noticias (Rodríguez & Ureña, 2012). Ni siquiera se necesita ya un ordenador, el uso a través de teléfonos móviles, permite la consulta en cualquier momento y en cualquier lugar. Twitter solo puede ser entendido por aquellos que lo han probado, adquiere pleno sentido cuando las relaciones entre personas pasan a adquirir una sensación de proximidad absoluta, incluso cuando se desarrollan a miles de kilómetros de distancia.

### **Es una fuente de información.**

Esta red se ha convertido para varios de sus usuarios en un medio a través del cual están informados, mucha gente reconoce que se entera de las noticias antes por Twitter que por cualquier otro medio de comunicación. De hecho, *“todas las noticias de alcance de los últimos años han aparecido inicialmente en Twitter”*. (Orihuela J. L., 2013)

Twitter es un cauce para generar noticias, los medios prestan mucha atención a todo lo que se dice en Twitter, produciéndose un efecto “altavoz” o “rebote”. Cada vez es más frecuente, ver noticias basadas en las “declaraciones” de un personaje a través de su cuenta de Twitter, esta dimensión pública de los tweets, diferente a la privacidad de los mensajes que caracteriza a otras redes sociales, como Facebook, hace que todo lo que se diga aquí sea transmitido como “si se hablara en voz alta para que todo el mundo lo oiga”. (Congosto, Fernández, & Moro Egido, 2011)

### **Ayuda a pensar y hablar en “titulares” y, por tanto, a ser mejores portavoces.**

Cualquier usuario puede trasladar sus mensajes siendo un narrador de hechos noticiosos de los que está siendo testigo a través de Twitter, en donde condensan las ideas principales que se quieren publicar, el hablar en “titulares”, frases breves, concisas y directas que sirvan para dar una información. Este medio puede servir como puerta de entrada a otras plataformas, puesto que los tweets pueden contener enlaces a páginas web, blogs, fotografías o vídeos donde las ideas condensadas en esos caracteres a modo de titular son desarrolladas con mayor profundidad. Y es que esta potencialidad es cada vez más empleada, porque se comprueba al ver cómo las redes sociales son ya, tras los motores de búsqueda, la principal fuente de tráfico de visitas hacia las páginas web. (Dans, 2011)

### **Humaniza a los usuarios y aumenta la empatía hacia ellos.**

Los personajes no solo se limitan a hablar sólo de un tema específico sino que dejan entrever opiniones y aspectos de su vida personal que acaban siendo del interés de los tuiteros. Contar con una buena imagen, que transmita confianza y con el que la gente se pueda identificar es un valor muy considerado, y por este motivo, dar a conocer sus hobbies, su infancia, aficiones y otros, exponen detalles de su pasado, sobre todo si es heroico o ejemplar, lo que ayuda a proyectar una sensación emocional de empatía.

Dar a conocer qué hacen los usuarios responde fielmente a la formulación de la pregunta que definía en sus orígenes a esta red social “¿What are you doing?” “¿Qué estás haciendo?”, esta visión se ajusta a una de las funciones de Twitter denominada “familiaridad ambiental”. (O’Reilly & Milstein, 2009)



### **Es un termómetro social.**

Indudablemente Twitter no puede sustituir en esta función a las empresas demoscópicas, al menos de momento, pero puede ofrecer también un conocimiento bastante certero de lo que piensa la gente sobre determinados temas, convirtiéndose así, en un “*gigantesco estudio de mercado en tiempo real y a escala global*” (Orihuela J. L., 2013) o, en “*el sensor de comportamientos sociales más certero de los estados de ánimo, las ideas reputacionales y los contextos públicos*”. (Gutiérrez-Rubí, 2014)

Y, de hecho, cada vez es más frecuente ver a personajes que tuitean propuestas o planes que tienen pensado ejecutar para que los usuarios les envíen sus puntos de vista y sugerencias, creando un gran brainstorming en el que contar con la participación de la opinión pública ayuda en la toma de decisiones.

### **Es un medio más en la estrategia global de comunicación.**

En el ámbito empresarial ya son muchos los que afirman que la presencia en Internet y, más específicamente en las redes sociales, ayuda a las compañías a aumentar sus ventas (Thomases, 2010), y quizá este argumento sea válido en la actualidad, pero lo que parece evidente es que nadie puede asegurar lo que depara un futuro en el que la población internauta aumente aún más y este medio acabe teniendo cotas de seguimiento similares a las que puede tener hoy día la televisión, la radio o la prensa. En definitiva, todo esto podría resumirse diciendo que, hoy por hoy, las herramientas 2.0 bien manejadas acercan a la gente y se posicionan potenciando un efecto viral.

#### **1.2.3. Modelo comunicativo de twitter.**

Según José Luis Orihuela estas son las características del modelo comunicativo de Twitter (Orihuela J. L., 2013):

- **Asimétrico:** Es una red social de relaciones optativas (seguir/ser seguido), en la que no se requiere el consentimiento mutuo entre los usuarios.
- **Breve:** Tiene un formato de escritura limitado a 140 caracteres por mensaje.
- **Descentralizado:** Posee una arquitectura variable multipunto-multipunto definida por las decisiones de cada usuario.
- **Global:** Provee un servicio disponible en varios idiomas y en todo el planeta.

- **Hipertextual:** Es un entorno de lecto-escritura en el que cada mensaje contiene enlaces por defecto (el usuario y el enlace permanente del mensaje) y en el que el uso del símbolo @ (arroba) y del signo # (numeral) genera enlaces de manera automática.
- **Intuitivo:** Propicia un concepto de aplicación y una interfaz web orientados a usuarios no expertos, basados en la simplicidad y la facilidad de uso.
- **Multiplataforma:** Facilita una aplicación con la que se puede interactuar desde clientes de mensajería, de correo, de SMS, navegadores web y sus extensiones, ordenadores de sobremesa, portátiles, netbooks, tablets, móviles y redes sociales.
- **Sincrónico:** Genera una temporalidad definida por la fugacidad del timeline en el que tienden a coincidir los tiempos de publicación y lectura.
- **Social:** Crea un conjunto de comunidades y relaciones definidas por cada usuario.
- **Viral:** Es una plataforma de carácter global, social y sincrónico, facilita la rápida circulación y multiplicación de los mensajes.

Y es que Twitter, más allá de la tecnología, son las personas que lo utilizan, los mensajes que se comparten, las conversaciones que se mantienen y las relaciones que se establecen entre sus usuarios.

#### 1.2.4. Metadatos de un tweet

Los metadatos son “datos sobre datos” o “información sobre los datos”, es decir, información estructurada que describe a otra información y que permite encontrarla, gestionarla, controlarla, entenderla y preservarla en el tiempo.

Los metadatos no son un concepto nuevo: la inclusión de metadatos es el método utilizado tradicionalmente en el mundo de las bibliotecas para gestionar libros y archivos. Sin embargo, el nacimiento de la informática y, más concretamente, las posibilidades de intercambio de información surgidas en torno a la Web Semántica (Classora, 2013), hizo que aumentara la popularidad del término.

Existen metadatos que se encuentran detrás de un mensaje del usuario, los mismos que describen y catalogan la información que contiene un tweet, no son visibles porque esa información se encuentra empaquetada en conjunto con el mensaje. (Jiménez , 2013). A continuación se indica los metadatos que contiene un tweet (Twitter Developers, 2013):

- **Contribuyentes**

Contiene información sobre el identificador único del usuario y el nombre de usuario que contribuyó a la autoría del tweet.

```
"contributors":  
[  
  {  
    "id":819797,  
    "id_str":"819797",  
    "screen_name":"episod"  
  }  
]
```

Figura 1: Contribuyentes de un tweet.

Fuente 1.-Tweet Developers

- **Coordenadas**

Presenta la ubicación geográfica del Tweet en su matriz de coordenadas: longitud y latitud, según lo informado por el usuario o la aplicación cliente.

```
"coordinates":  
{  
  "coordinates":  
  [  
    -75.14310264,  
    40.05701649  
  ],  
  "type":"Point"  
}
```

Figura 2: Coordenadas de un tweet

Fuente 2.- Tweet Developers

- **Fecha de Creación**

Indica la fecha y hora en el que el tweet fue creado de acuerdo al UTC (Tiempo Universal Coordinado).

```
"created_at":"Wed Aug 27 13:08:45 +0000 2008"
```

Figura 3: Fecha de creación de un tweet

Fuente 3.- Tweet Developers

- **Lenguaje**

Muestra un identificador de idioma correspondiente al lenguaje detectado por la máquina en el texto del Tweet.

```
"lang": "en"
```

Figura 4: Lenguaje de un tweet

Fuente 4.- Tweet Developers

- **Favorito**

Indica si el Tweet es favorito por el usuario que se autentica.

```
"favorited":true
```

Figura 5: Tweet favorito

Fuente 5. - Tweet Developers

- **Número de favoritos**

Indica aproximadamente cuántas veces el Tweet es marcado como favorito por los usuarios de Twitter.

```
"favorite_count":1138
```

Figura 6: Número de favoritos de un tweet

Fuente 6.- Tweet Developers

- **Retweet**

Indica si el Tweet ha sido retwitteado por el usuario que se autentica.

```
"retweeted":false
```

Figura 7: Tweet retwitteado

Fuente 7. - Tweet Developers

- **Número de Retweet**

Presenta el número de veces que un tweet ha sido retweeted.

```
"retweet_count":1585
```

Figura 8: Número de retweets

Fuente 8.- Tweet Developers

- **Lugar**

Contiene información del lugar donde fue publicado el tweet con sus correspondientes coordenadas geográficas, país, código, calles, tipo de lugar, y su respectiva url que da a conocer más del lugar.

```
"place":
{
  "attributes": {},
  "bounding_box":
  {
    "coordinates":
    [
      [
        [-77.119759, 38.791645],
        [-76.909393, 38.791645],
        [-76.909393, 38.995548],
        [-77.119759, 38.995548]
      ],
      "type": "Polygon"
    ],
    "country": "United States",
    "country_code": "US",
    "full_name": "Washington, DC",
    "id": "01f8e706f872cb32",
    "name": "Washington",
    "place_type": "city",
    "url": "http://api.twitter.com/1/geo/
/id/01f8e706f872cb32.json"
  }
}
```

Figura 9: Lugar de un tweet

Fuente 9.- Tweet Developers

- **Entidades**

Las entidades son objetos que se encuentran dentro del contenido del tweet.

```
"entities":
{
  "hashtags": [],
  "urls": [],
  "user_mentions": []
}
```

Figura 10: Entidades de un tweet

Fuente 10.-Tweet Developers

- **Hashtags**

Representa las etiquetas que se encuentran dentro de un tweet, en la matriz de enteros se indican los desplazamientos de donde comienza y termina una etiqueta, y el nombre de la misma.

```
"hashtags": [{"indices": [32, 36], "text": "lol"}]
```

Figura 11: Hashtags de un tweet

Fuente 11.-Tweet Developers

- **Urls**

Representa urls incluidas en el texto de un Tweet, display\_url indica la url corta y expanded\_url indica la url extendida.

```
"urls": [{"indices": [32, 52], "url": "http://t.co/\\IOwBrTZR", "display_url": "youtube.com/watch?v=oHg5SJ\\u2026", "expanded_url": "http://www.youtube.com/watch?v=oHg5SJYRHA0"}]
```

Figura 12: Urls de un tweet

Fuente 12.-Tweet Developers

- **User\_mentions**

Representa a otros usuarios de Twitter que se mencionan en el texto del Tweet.

```
"user_mentions": [{"name": "Twitter API", "indices": [4, 15], "screen_name": "twitterapi", "id": 6253282, "id_str": "6253282"}]
```

Figura 13: Menciones de un tweet

Fuente 13.-Tweet Developers

- **Media**

Representa elementos multimedia cargados con el Tweet.

```
"media": [{"type": "photo", "sizes": {"thumb": {"h": 150, "resize": "crop", "w": 150}, "large": {"h": 238, "resize": "fit", "w": 226}, "medium": {"h": 238, "resize": "fit", "w": 226}, "small": {"h": 238, "resize": "fit", "w": 226}}, "indices": [15, 35], "url": "http://t.co/rJC5Pxsu", "media_url": "http://p.twimg.com/AZVLmp-CIAAbkyy.jpg", "display_url": "pic.twitter.com/rJC5Pxsu", "id": 114080493040967680, "id_str": "114080493040967680", "expanded_url": "http://twitter.com/yunorno/status/114080493036773378/photo/1", "media_url_https": "https://p.twimg.com/AZVLmp-CIAAbkyy.jpg"}]
```

Figura 14: Elementos multimedia de un tweet

Fuente 14.-Tweet Developers

- **Texto**

Indica el texto de una actualización de estado en formato UTF-8.

```
"text": "Tweet Button, Follow Button, and Web  
Intents javascript now support SSL http:\\\\t.co  
\\9fbA0oYy ^TS"
```

Figura 15: Texto de un tweet

Fuente 15.-Tweet Developers

- **Usuario**

Posee información del usuario que publicó un tweet, da a conocer el número de tweets, número de tweets favoritos, color de perfil, nombre de usuario, número de seguidores y otros.

```
"user": {"statuses_count": 3080,  
"favourites_count": 22, "protected": false,  
"profile_text_color": "437792",  
"profile_image_url": "...", "name": "Twitter API",  
"profile_sidebar_fill_color": "a9d9f1",  
"listed_count": 9252, "following": true,  
"profile_background_tile": false,  
"utc_offset": -28800, "description": "The Real  
Twitter API. I tweet about API changes, service  
issues and happily answer questions about Twitter  
and our API. Don't get an answer? It's on my  
website.", "location": "San Francisco, CA",  
"contributors_enabled": true, "verified": true,  
"profile_link_color": "0094C2",  
"followers_count": 665829, "url": "http:\\/  
\\dev.twitter.com", "default_profile": false,  
"profile_sidebar_border_color": "0094C2",  
"screen_name": "twitterapi",  
"default_profile_image": false,  
"notifications": false, "display_url": null,  
"show_all_inline_media": false, "geo_enabled": true,  
"profile_use_background_image": true,  
"friends_count": 32, "id_str": "6253282", "entities":  
{ "hashtags": [], "urls": [], "user_mentions": [] },  
"expanded_url": null, "is_translator": false,  
"lang": "en", "time_zone": "Pacific Time (US & Canada)",  
"created_at": "Wed May 23 06:01:13 +0000  
2007", "profile_background_color": "e8f2f7",  
"id": 6253282, "follow_request_sent": false,  
"profile_background_image_url_https": "...",  
"profile_background_image_url": "...",  
"profile_image_url_https": "..."}
```

Figura 16: Usuario de twitter

Fuente 16.-Tweet Developers

### 1.2.5. Tipos de metadatos

En general, los metadatos pueden clasificarse en tres amplias categorías con límites no siempre bien definidos, y muchas veces superpuestos (Classora, 2013):

**Metadatos descriptivos:** Son aquellos que sirven para la descripción e identificación de los recursos de información. Permiten la búsqueda y recuperación de los datos, así como distinguir recursos y entender sus contenidos. Se realizan mediante estándares como Dublin Core, estándares MARC, microdatos de HTML y otros.

**Metadatos estructurales:** Son los que más influyen en la recuperación de la información, facilitando la navegación y presentación de los recursos. Ofrecen información sobre su estructura interna y permiten establecer relaciones entre ellos, de manera que pueden unir los archivos de imagen y textos que están relacionados. Los estándares más difundidos son: SGML, RDF y EAD (Encoded Archival Description).

**Metadatos administrativos:** Incluyen datos para la gestión de derechos, firma digital, auditorías de calidad, control de acceso mediante contraseñas. También se podrían enmarcar en esta categoría los metadatos sobre preservación a largo y corto plazo. Ejemplo, imágenes DICOM que incluyen metadatos como: tipo y modelo de escáner utilizado, resolución, paciente, limitaciones de reproducción. Los estándares más utilizados son: CopyrightMD, DOI (Digital Object Identifier), ODRL (Open Digital Rights Language).

### 1.2.6. Datos dentro del contenido de un tweet

La información que se puede encontrar dentro del contenido de un tweet (Suh, Hong, Pirolli, & Ed, 2010) , se detalla a continuación:

- ***Etiquetas / Hashtags.***

Un hashtag es una convención adoptada por los usuarios de Twitter para relacionar sus mensajes con un tema específico (Carter, Tsagkias, & Weerkamp, 2011). El hashtag o etiqueta se compone de una cadena de caracteres corta precedida del símbolo “#”.



Twitter estructura la información mediante el hashtag, que lo utiliza para marcar palabras clave en un Tweet que hace referencia a un tópico relevante. Fue un recurso creado orgánicamente por la comunidad de usuarios de Twitter, como una manera de categorizar los mensajes.

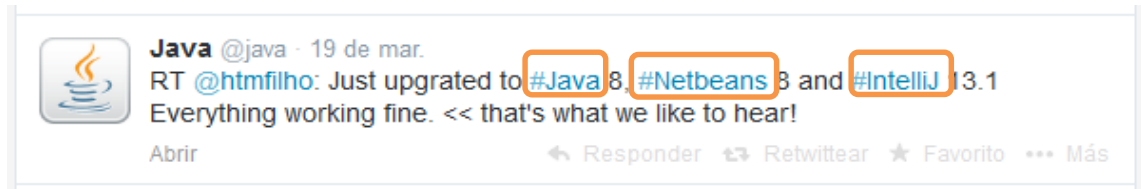


Figura 17: Etiquetas de un tweet

De la misma manera los hashtags permiten crear un canal de comunicación paralelo a nuestra línea del tiempo, permitiendo reunir información relativa a un tema concreto sin necesidad de seguir a los usuarios que lanzan mensajes con esa etiqueta, su divulgación y dispersión depende de la red de los seguidores que deseen participar en la conversación.

- **Menciones.**

Una mención es cualquier actualización de Twitter que contenga “@nombredeusuario” en el cuerpo del Tweet, sin importar en que parte del mensaje está, lo que significa que las @respuestas también son consideradas menciones. (Twitter, 2013)

Las menciones, no siempre le hablan al usuario sino que en ocasiones se usa para referir a un usuario, buscando que la persona tenga conocimiento de esto. Los tweets que incluyen menciones son empleados para informar al destinatario que se ha publicado un mensaje que está dirigido a él, se trata de un mensaje, en principio público y visible por cualquiera, que en el perfil del usuario destinatario irá marcado de forma especial para darle prioridad. (Collado C, 2011)



Figura 18: Mención de un tweet

- **Texto.**

Es útil que el tweet contenga texto<sup>5</sup>, que es un entramado de signos con una intención comunicativa que busca transmitir un cierto mensaje que adquiere sentido de acuerdo a cada contexto, pudiendo ser una idea, opinión o conversación acerca de algún asunto o tema en específico. Se puede aislar las menciones, hashtag y url, para tomar solo el texto.

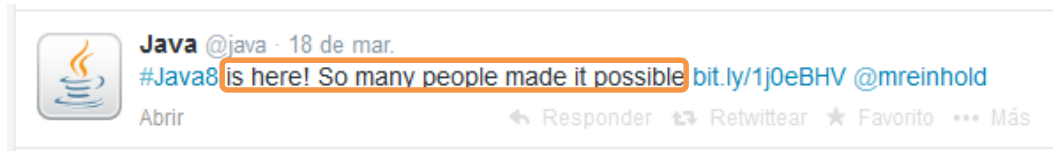


Figura 19: Texto de un tweet

- **Links.**

Las URL son un Localizador Uniforme de Recursos (URL por sus siglas en Inglés) es un sitio web que apunta a una página única en la Internet. (Twitter, 2013)

Se ha demostrado que los tweets más retweets son los que comparten enlaces o links en su contenido (Twitter, 2013). Las personas se interesan más por tweets que contienen una url con información adicional.



Figura 20: Link de un tweet

- **URL acortada.**

El acortamiento de URL es una técnica en la World Wide Web en el que un localizador uniforme de recursos (URL) podrá hacerse sustancialmente más corto en longitud y sigue directamente a

---

<sup>5</sup> Obtenido de <http://es.wikipedia.org/wiki/Texto>

la página deseada, esto se logra mediante el uso de una redirección HTTP en un nombre de dominio que sea corto, que conecta a la página web que tiene una URL extendida.

Esto es conveniente para las tecnologías de mensajería tales como Twitter e Identi.ca que limitan severamente el número de caracteres que pueden utilizarse en un mensaje. Las Urls cortas permiten hacer referencia a direcciones web largas que son indicadas dentro un tweet, donde es más fácil encajar las URL largas en el límite de los 140 caracteres que permite Twitter.

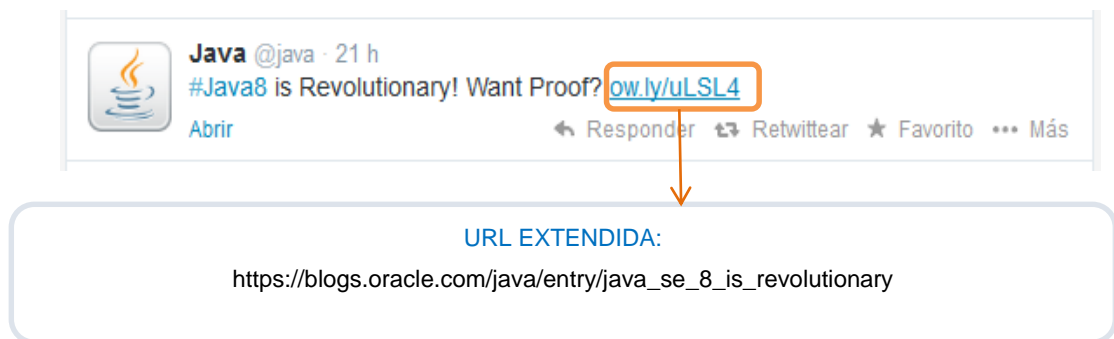


Figura 21: Url acortada de un tweet

### 1.2.7. Estructura del tweet

La estructura indica el conjunto de todos los elementos importantes que forman un mensaje, se presentan básicamente dos estructuras de tweets que se las ha categorizado de la siguiente manera:

- **Estructura simple:** Consiste en un tweet formado por un texto de 140 caracteres o menos.

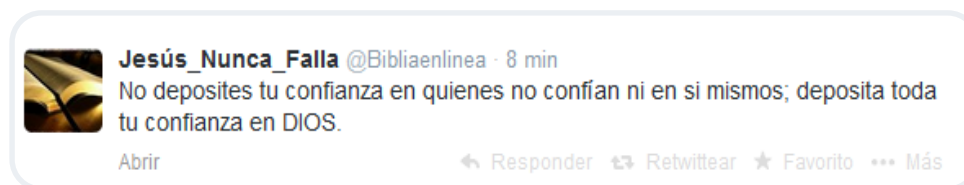


Figura 22: Estructura simple de un tweet

- **Estructura compuesta:** A diferencia de los tweets de estructura simple, los de estructura compuesta se forman porque además del texto incluyen #etiquetas,

@menciones, /via@nombreUsuario retwitteado, enlace acortado, obteniendo un tweet con sus todos sus elementos.

No afecta si tiene una o más etiquetas, menciones, urls, texto, siempre y cuando no supere el límite de los caracteres que propicia la red social Twitter.



Figura 23: Estructura compuesta de un tweet

A continuación se muestran las convenciones en la estructura de un tweet, los elementos que lo conforman pueden ir en cualquier parte del mensaje:

Tweet para interactuar, compartir enlaces, y etiquetas:

- @Nombre del usuario + texto: Interactúa con otro usuario.
- Texto + enlace: Comparte enlaces.
- Texto + #hashtag: Comparte etiquetas.
- @Nombre del usuario + texto + #hashtag+ enlace: Interactúa, comparte enlaces y etiquetas.

Para hacer RT a un tweet:

- RT + @nombre de usuario retwitteado + Tweet
- Tweet + / vía @nombre del usuario retwitteado

### 1.2.8. Indicadores twitter

Las redes sociales como Twitter parten principalmente de los motores de búsqueda, pero arrastran influencias de los medios de comunicación tradicionales y así se habla de ciertos indicadores (Orta, Pardo, & Salas, 2012), como:

- Seguidores: Es el número total de seguidores del usuario, lo que da una estimación de los que directamente visualizarán el mensaje que un usuario emita y por lo tanto puede ser una primera medida directa de su influencia, aunque no toma en cuenta más características de la subred del usuario.
- Influencia de red: Este valor no sólo tiene en cuenta los seguidores sino su influencia, de manera recursiva se calcula el número de seguidores que tienen y la influencia de estos.
- Amplificación de red: Muestra la influencia directa que tiene un usuario sobre su red, es decir, cuando emite un mensaje, cuál es la probabilidad de que ese mensaje sea retransmitido y por lo tanto amplificado.
- Alcance real: Es un indicador que a partir de la influencia de red y la amplificación de red calcula el valor de audiencia real a la que puede llegar el usuario con un mensaje.

Estas tres actividades representan los diferentes tipos de influencia de una persona (Cha, Haddadi, Benevenuto, & Gummadi, 2010):

- 1) Influencia grado de entrada, es el número de seguidores de un usuario que indica directamente el tamaño de su audiencia.
- 2) Influencia Retweet, se mide a través del número de retweets que contengan el nombre de uno, indica la capacidad de ese usuario para generar contenido con paso de largo el valor.
- 3) Mención de influencia, que se mide a través del número de menciones que contiene el nombre de uno, indica la capacidad que tiene dicho usuario para involucrar a otros en una conversación.

### **1.3. Análisis de redes sociales**

Las Redes Sociales son herramientas útiles que permiten informar, comunicar y conectar con personas en cualquier parte del mundo. El acceso al Internet es diferente en cada región, sin embargo existe una necesidad latente del ser humano por interactuar con otros que han hecho que esas brechas digitales sean cada vez menores. (iLifebelt, 2013)

En sentido amplio, una red social es una estructura social formada por personas o entidades conectadas y unidas entre sí por algún tipo de relación o interés común. Las mismas son parte de nuestra vida, puesto que son la forma en la que se estructuran las relaciones personales, estando conectados mucho antes de tener conexión a Internet. (Ponce, 2012)

El análisis de las redes sociales también ha sido llevado a cabo por especialidades como la estadística, el marketing, la publicidad, puesto que la mejor forma de conseguir una muestra poblacional diversa y variada es a través de internet, estas ramas suelen hacer estudios de mercado, datos, investigación o de análisis para determinar gustos, patrones, crecimientos poblacionales, entre otros factores.

También las matemáticas y las ciencias de la computación realizan estudios basados en redes sociales, aplicando técnicas, métodos, algoritmos y teorías, sin entrar en más detalle, la teoría de los grafos representa las redes sociales (Mejia, 2010) mediante nodos conectados por aristas, donde los nodos serían los individuos y las aristas las relaciones que los unen. Todo ello conforma un grafo, una estructura de datos que permite describir las propiedades de una red social.

En el análisis de redes se describen y estudian las estructuras relacionales que surgen cuando diferentes organizaciones o individuos interactúan, se comunican, coinciden y colaboran, a través de diversos procesos o acuerdos, que pueden ser bilaterales o multilaterales (Lugo, 2011); de este modo la estructura que surge de la interrelación se traduce en la existencia de una red social.

Entonces al igual que Sanz se puede concluir que: “Las redes sociales son, por tanto, conjuntos de relaciones sociales o interpersonales que ligan individuos u organizaciones en grupos.” (Sanz, 2003)

Se demuestra que las redes sociales juegan un papel crucial porque se han convertido en un factor crítico en la influencia, búsqueda, y difusión de la información, en donde la estructura de la red afecta a la dinámica del flujo de información.

Los investigadores tienen acceso a grandes cantidades de datos para su análisis, estos conjuntos de datos ofrecen una rica fuente de evidencia para el estudio de la dinámica de la persona y comportamiento del grupo, la estructura de las redes y los patrones globales del flujo de información sobre ellos. (Lerman & Ghosh, 2010)

#### **1.4. Problemática**

La actividad del usuario en las redes sociales en línea se está expandiendo. Twitter, da acogida todos los días a la actividad del intercambio de información de miles de millones de los usuarios, el uso de esta red social permite que las personas se comuniquen ideas, opiniones, vídeos y fotos entre sus círculos de amigos y seguidores en todo el mundo.

Estas interacciones generan una gran cantidad de datos sin precedentes que pueden ser utilizados para realizar un análisis de redes sociales que proporcionan una oportunidad para arrojar luz sobre los mecanismos de la difusión de información a través de la comunicación humana, esta información fluye de manera viral a partir de la creación y generación de un tweet y su posterior redifusión por la red hasta que el mensaje muere. Sin embargo, no se conoce cómo el comportamiento de la información en línea puede haber cambiado, es por este motivo que se pretende estudiar los cambios en las etiquetas, menciones, urls y texto de todos los tweets que comparten una misma url en la red social twitter.

#### **1.5. Solución**

Mediante la metodología de análisis de redes sociales la solución es aplicar métricas que permitan determinar los cambios que existen en la descripción de recursos publicados como links en la red social twitter, es decir, tweets que comparten una misma url extendida. Estos cambios serán medidos a través de «hashtags», «menciones», «urls» y «texto». Y para dar cumplimiento al mismo se propone diseñar e implementar un algoritmo que mida el número de cambios que experimenta un recurso en cada una de sus métricas, para finalmente poder realizar un estudio real de la difusión y re-difusión de información en Twitter.

## **1.6. Trabajos relacionados con la difusión de información**

### **Contagio de la influencia basada en la difusión de homofilia impulsada en redes dinámicas.**

La difusión depende del tipo de contagio. Contagios complejos, a diferencia de las enfermedades infecciosas (contagios simples), se ven afectados por el refuerzo social y homofilia. Por lo tanto, la propagación en comunidades altamente clúster es mayor, mientras que la difusión a través de las comunidades se ve obstaculizada. Una hipótesis común es que los memes y comportamientos son contagios complejos. Se demuestra que, si bien la mayoría de los memes de hecho se han extendido como contagios complejos, algunos memes virales se propagan a través de muchas comunidades, como las enfermedades. Las características de nodo y comportamientos a menudo se relacionan con la estructura de las redes sociales a través del tiempo. (Weng, Lilian; Menczer, Filippo; Ahn, Yong Yeol, 2013)

### **Diferencias en la mecánica de difusión de la información.**

El trabajo trata sobre la “Variación en la Propagación de Hashtags”. Donde se estudian las fuentes de variación mediante la simulación de modelos matemáticos que permite analizar la forma de que los hashtags más utilizados en Twitter se propagan dentro de su población de usuarios. Se encuentra que estas fuentes de variación no sólo implican diferencias en la probabilidad con la que algo se contagia de una persona a otra. (Romero, Meeder, & Kleinberg, 2011)

### **Impacto de los patrones de actividad humana sobre la dinámica de difusión de la información.**

Estudia el impacto de los patrones de la actividad humana sobre la difusión de información. Con un experimento de correo electrónico viral que incluyó 31.183 personas en el que han sido capaces de rastrear una pieza específica de información a través de la red social. Se ha encontrado que la información viaja a un ritmo lento inesperadamente. Mediante el uso de un modelo de ramificación que describe con precisión el experimento, se muestra que la gran heterogeneidad que se encuentra en el tiempo de respuesta es el responsable de la lenta dinámica de la información a nivel colectivo. (Iribarren, 2009)



## 1.7. Trabajos relacionados con twitter.

### Difusión de la información en twitter.

Presenta dos tipos de difusión: "simple" vs "complejo", en donde los simples necesitan sólo una persona y los complejos necesitan varias personas, estrictamente se enfocan en un modelo simple de la difusión de información, en el cual realizan un experimento para analizar la difusión de un tema en twitter. Recopilan datos únicos y populares apuntados a las tendencias de esta red, como los hashtag que se propagan de persona a persona, y evalúan el mecanismo de difusión utilizando tweets y retweets con etiquetas, para entender cómo se propaga un tema, el comportamiento que tuvo, saber porque unos temas se propagan y otros no, se puede detectar cuando un tópico se va a difundir dentro de una comunidad o un grupo.

Para filtrar los tweets usaron expresiones regulares que coincida con cualquiera de los temas elegidos, manejan la difusión como la toma de decisiones en donde la información local debe elegir umbrales que significan el comportamiento, basado en lo que los vecinos están haciendo. Como resultados se obtuvo que los clúster son un obstáculo para la difusión de información generalizada (en teoría), que el seguimiento de la difusión de información en una red real es no trivial, haciendo hincapié en los modelos simples ya que pueden ser útiles conceptualmente pero difícil de aplicar a los datos reales. (Nikolov, 2012)

### Información contagio: Estudio empírico de la propagación de noticias en las redes sociales Digg y Twitter.

Este trabajo presenta un estudio empírico de la actividad del usuario sobre la difusión de información en las redes sociales Digg<sup>6</sup> y Twitter<sup>7</sup>. Para el estudio se recogieron datos sobre las historias populares de estos sitios, que incluye información acerca de quién votó o retweeted la historia y cuando lo hizo. Además, se extrajo usuarios activos de estas redes para dar seguimiento de cómo el interés en las noticias se extiende entre ellos.

Para estudiar cómo la información se propaga a través de la red, primeramente se identifica la estructura de las redes sociales en ambos sitios, luego se mide la actividad del usuario en la evolución del número de votos que una historia recibe, utilizando la similitud entre los individuos por distancia como un número de bordes que separan dos nodos dentro de un gráfico.

---

<sup>6</sup> <http://digg.com/>

<sup>7</sup> <https://twitter.com/>

Aunque los dos sitios son muy diferentes en su funcionalidad y la interfaz de usuario, utilizan formas similares de difundir información. En los resultados se demuestra que la estructura de la red afecta a la dinámica de flujo de información, la información llega a los nodos más rápido en una red más densa como Digg, mientras que en Twitter las historias se propagan más lejos. Por otra parte, el alcance de la información que se propaga no parece depender de la similitud entre los usuarios, por lo menos cuando la similitud se mide por el número de bordes entre ellos. En Digg, aunque sus usuarios están altamente interconectados, una historia no llega a la mayor cantidad de fans como en Twitter, donde los usuarios se conectan con menor densidad. Sin embargo, el número de votos acumulados por las historias de ambos sitios satura después de un período de alrededor de un día a un valor que refleja su popularidad. (Lerman & Ghosh, 2010)

### **Un procedimiento de difusión de la información.**

Se presenta un procedimiento de diseminación de la información en una red general. El criterio utilizado para hacer eficiente la elección de las ramas del emparejamiento en cada etapa del proceso consiste en asociar a cada rama un peso determinado por la medida de equilibrio del subconjunto de vértices no informados, aprovechando que esta medida tiene información sobre la conectividad de este subconjunto con su complementario. Describen la aplicación del algoritmo que permite realizar el proceso de difusión de la información en una red general, en los resultados comprueban que el algoritmo propuesto para el proceso de difusión de la información no aplica a redes con un gran número de vértices debido a que no utilizan métodos iterativos en la medida de equilibrio. (Amblás, Bendito, Carmona, Encinas, Joan, & Mitjana, 2003)

## **CAPÍTULO 2**

### **DISEÑO E IMPLEMENTACIÓN DEL ALGORITMO**

Al resolver problemas de distinta índole, en muchas ocasiones se sigue una metodología para conseguir tal propósito. Dicha metodología se encuentra caracterizada por una serie de acciones o situaciones llevadas a cabo (Vazquez Gomez, 2012). Las cuales tienen como propósito, cada una de ellas lograr un objetivo en específico y en conjunto alcanzar un objetivo general. En consecuencia, estas acciones se pueden concebir como algoritmos.

Por esto se presenta una alternativa de solución mediante el desarrollo de un algoritmo que determine los cambios de los recursos publicados en Twitter, el lenguaje de programación como la computadora son los medios para obtener un fin: conseguir que el algoritmo se ejecute y se efectúe el proceso correspondiente.

## **2.1. Representación de algoritmos**

Es preciso disponer de alguna notación algorítmica que permita expresar la resolución de un algoritmo, estas notaciones deben ser independientes de modo que las sucesivas acciones no dependan de la sintaxis del lenguaje a utilizar en la fase de implementación, sino que su descripción pueda servir fácilmente para la transformación en un programa, es decir su codificación. Las notaciones algorítmicas más usuales son:

### **2.1.1. Diagrama de flujo**

Representa la esquematización gráfica de un algoritmo, que indica los pasos o procesos en orden lógico. Se basan en la utilización de diferentes símbolos para representar operaciones específicas, que se conectan por medio de flechas para indicar la secuencia de cada operación (Millán, Fermín, & Chacón, 2008). Su correcta construcción es sumamente importante porque si el diagrama de flujo está completo y correcto, el paso del mismo a un lenguaje de programación es relativamente simple y directo.

### **2.1.2. Diagrama Nassi-Shneiderman**

Es una representación gráfica de un algoritmo para programación estructurada, el problema es reducido en subproblemas cada vez menores, representadas en cajas anidadas que reflejan una descomposición de una forma clara y simple, hasta que solo comandos y estructuras de control permanecen, esto hace que los procesos del algoritmo sean más fáciles de representar y de interpretar. Este diagrama combina la descripción textual, propia del pseudocódigo, con la

representación con la representación gráfica del diagrama de flujo (Pérez Pérez & Monsalve López, 2012), además se apoya en un conjunto de palabras reservadas. (Juárez, 2009)

### 2.1.3. Pseudocódigo

Es una serie de normas léxicas y gramaticales como los lenguajes de programación, pero sin llegar a la rigidez de sintaxis de los mismos, el pseudocódigo es un lenguaje artificial e informal útil para el desarrollo de algoritmos, su objetivo es permitir que el programador se centre en los aspectos lógicos de la solución a un problema. Está considerado como una descripción de un algoritmo que resulta independiente de otros lenguajes de programación, para que pueda ser leído por un ser humano y no interpretado por una máquina.

## 2.2. Comparativa de las notaciones algorítmicas

A continuación se presenta las ventajas e inconvenientes de cada una de las notaciones para la representación de algoritmos ( Rosado Hernández, 2013):

Tabla 1: Notaciones algorítmicas para la representación de algoritmos

NOTACIÓN	VENTAJAS	INCOVENIENTES
<b>Diagramas de flujo</b>	Al ser visuales son muy sencillos de entender y para su diseño se utilizan símbolos estándar.	Dificultad de mantenimiento y actualización, puesto que debe utilizar editores gráficos.
<b>Diagrama de Nassi-Schneiderman</b>	Adecuación a la programación estructurada, favorecen las técnicas de diseño descendente, reduce los programas en módulos pequeños; resaltando más las partes generales sobre los detalles concretos, que quedan más pequeños cuantos más específicos son.	Dificultad de modificación.
<b>Pseudocódigo</b>	Facilidad de creación, evolución y mantenimiento, y la facilidad para expresar el pseudocódigo en cualquier lenguaje de programación.	Falta de estandarización y la dificultad para su lectura cuando su tamaño crece.

Según lo expuesto, la notación seleccionada para representar de manera general el algoritmo que determine los cambios es el diagrama de flujo.

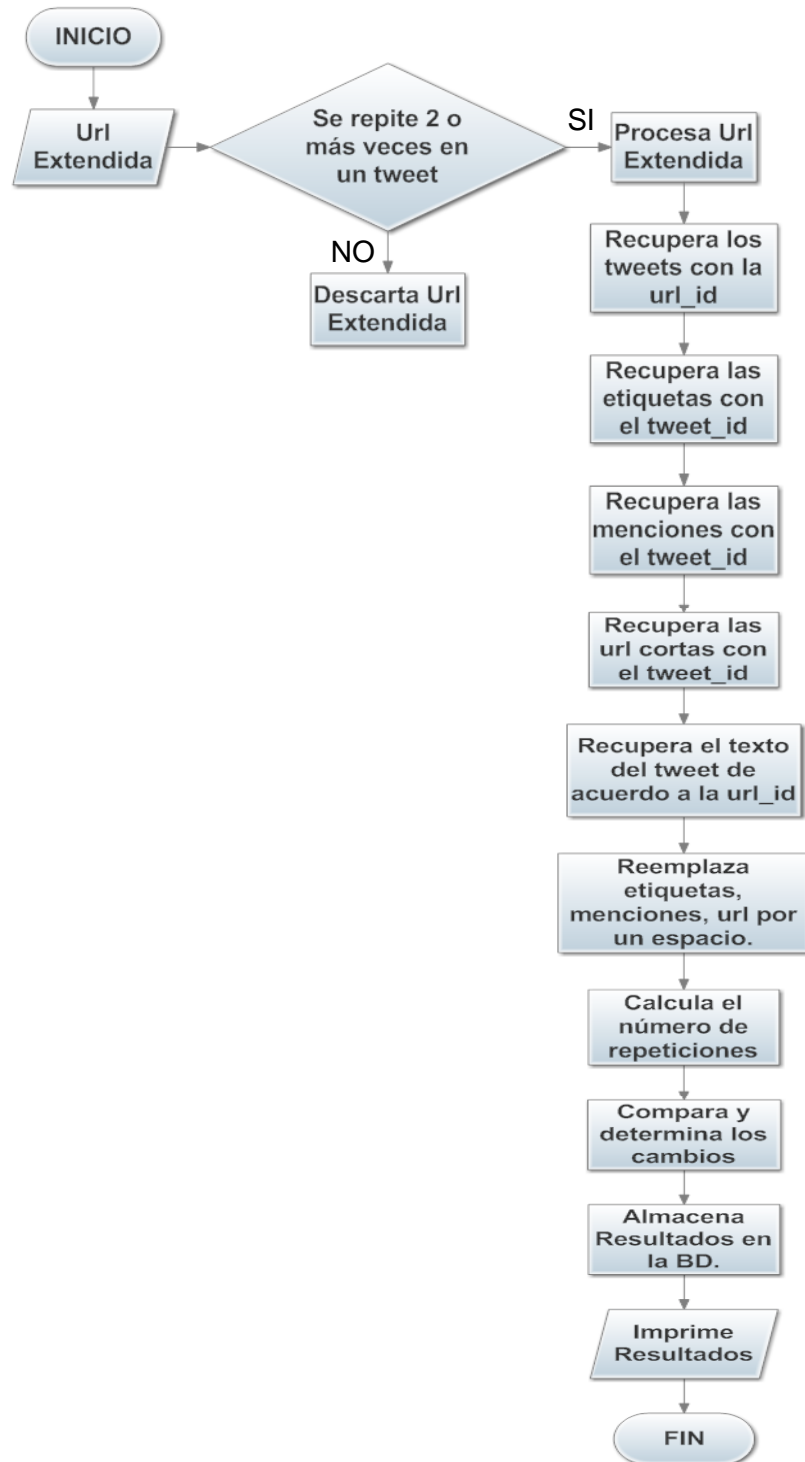


Figura 24: Diagrama de flujo del algoritmo

### **2.3. Algoritmo**

Está implementado en el lenguaje de programación java en el entorno de desarrollo netbeans, el código se encuentra almacenado en el repositorio de gitlab<sup>8</sup>, el propósito del algoritmo es medir los cambios que experimenta un recurso, para ello es importante hacer un seguimiento de cada tweet mediante los atributos que desencadena cada mensaje. El dominio de análisis va orientado al tema de urls extendidas, donde el punto de unión son los tweets que comparten una misma url, por lo que es preciso encontrar las similitudes entre los atributos de cada tweet (Ozdikis, Senkul, & Oguztuzun, 2012), de manera que se ejecutará el algoritmo sobre un grupo de datos.

#### **2.3.1. Entorno de desarrollo NetBeans.**

Se trabajó con NetBeans IDE 7.2.1 que es un entorno de desarrollo integrado galardonado disponible para Windows, Mac, Linux y Solaris. El proyecto NetBeans consiste en un IDE de código abierto “open source” que permite a los desarrolladores crear aplicaciones de escritorio, móviles, web utilizando la plataforma Java; así como PHP, JavaScript, Ajax, Groovy y Grails, y C / C+ +.

Se lo escogió porque ofrece soporte y es diseñado para el desarrollo de aplicaciones haciendo uso de la tecnología Java, proporcionando una arquitectura de aplicación fiable y flexible, siendo un producto libre y gratuito sin restricciones de uso ya que tiene una gran comunidad de usuarios y desarrolladores de todo el mundo.

#### **2.3.2. Servidor de base de datos**

SQLyog es una herramienta gráfica fácil de usar, compacta y muy rápida que permite administrar una base de datos MySQL, ofrece soporte para funciones en SQL y soporta diversas operaciones sobre la base de datos, dispone de una conexión jdbc para aplicaciones en java.

SQLyog es el gestor más potente de MySQL una herramienta de administración, que combina las características de MySQL Administrador, phpMyAdmin y herramientas visuales de MySQL. (Webyog, 2014)

---

<sup>8</sup> <http://hidrogeno.utpl.edu.ec:8084/gmagila/difusion-de-la-informacion-en-twitter/tree/master>

### 2.3.3. Descripción del proceso del algoritmo

Recordando que el punto de unión de los tweets son las URL extendidas, se analiza y mide los cambios de los tweets que publican una misma URL extendida, determinando el número de cambios de acuerdo a las siguientes métricas:

- Etiquetas
- Menciones
- Urls
- Texto

A continuación se ha establecido algunas fases que se encuentran debidamente ordenadas para poder determinar el proceso que mida los cambios de las métricas planteadas, cuyos resultados sirvan finalmente para realizar un estudio real de los cambios obtenidos.

Fases para determinar los cambios.

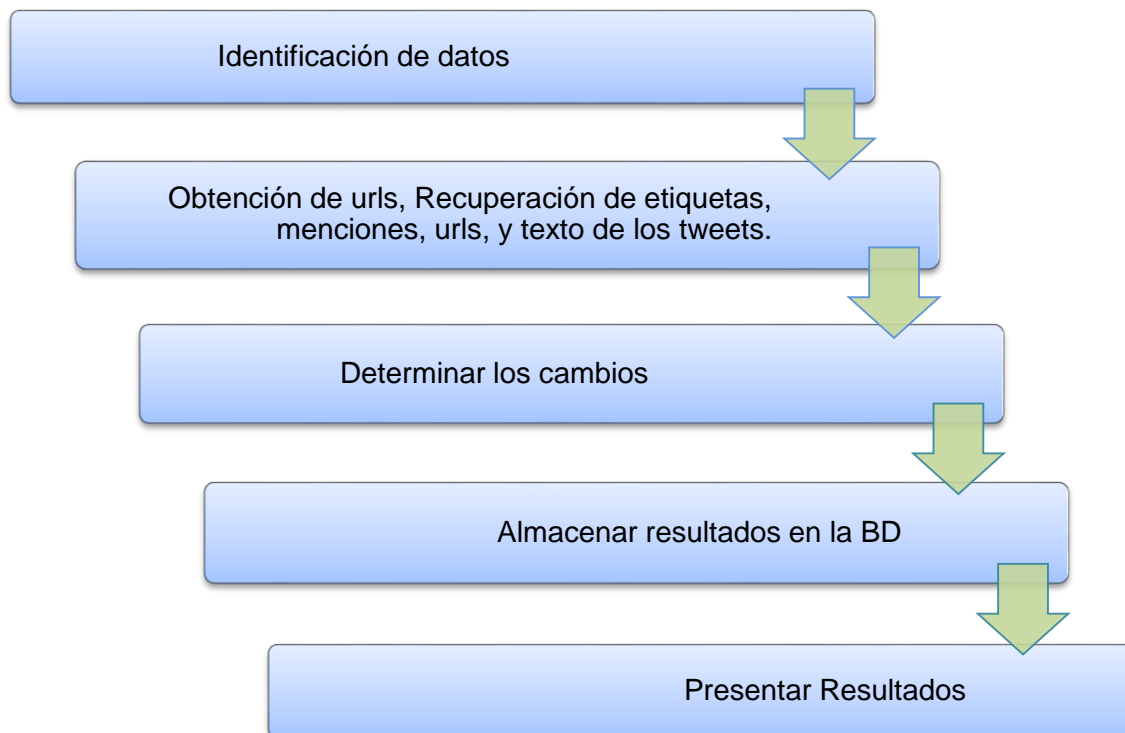


Figura 25: Proceso general para determinar los cambios



A continuación se detalla las tareas que comprende cada una de las Fases del Proceso para determinar los cambios de los tweets que comparten una misma url, en el diagrama de clases<sup>9</sup> se describe la estructura de la aplicación, indicando las clases con sus relaciones y los métodos que cumplen las tareas indicadas.

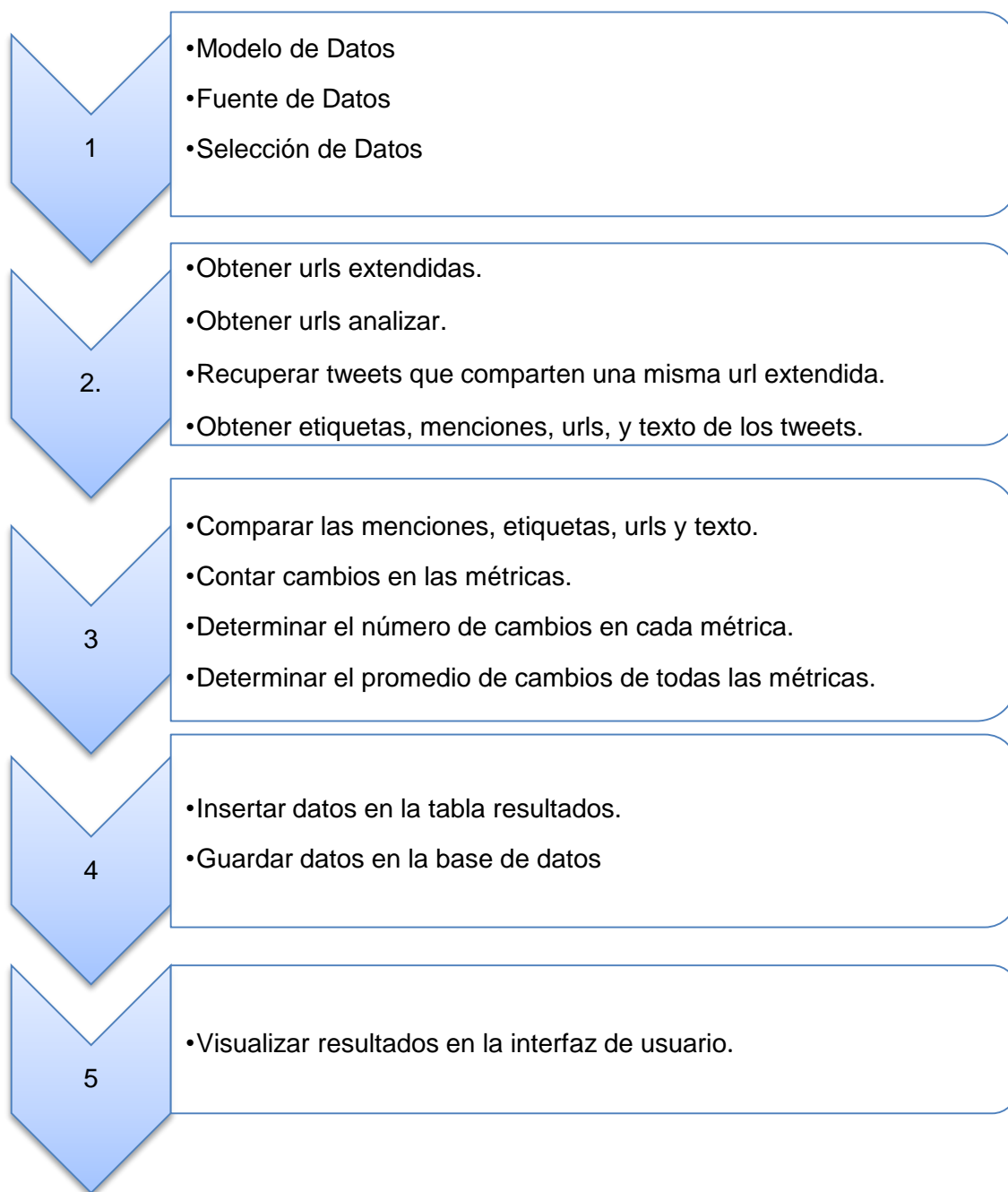


Figura 26: Tareas del las fases del proceso de cambios.

<sup>9</sup> Ver anexo 1.1.

### 2.3.3.1. Identificación de datos

Se utiliza la base de datos SNA\_TWITTER, que se encuentra almacenada en MySQL, se la puede apreciar en el modelo relacional de la base de datos.<sup>10</sup> Contiene un registro de 315252 tweets, 21590 etiquetas, 192186 usuarios y 61913 urls, su estructura consta de 11 tablas, siendo 7 de estas las que almacenan los datos requeridos para el desarrollo del proceso que determine los cambios.

Tabla 2: Datos para el estudio de los cambios

TABLA	CONTIENE
twett	información del tweet
url_t	urls extendidas
linked	urls acortadas
tag_t	etiquetas
user_t	usuario mencionado
tagged	datos intermedios
mention	datos intermedios

### 2.3.3.2. Obtención de urls

- **Urls extendidas**

Primero se recupera todas las urls extendidas que serán analizadas, se encuentran almacenadas en la tabla “url\_t” en donde se trabaja con los atributos “URL\_ID” con su respectiva url extendida que en este caso es el atributo “EXTENDED”.

Se obtiene un total de 61913 registros de urls extendidas, en la imagen se muestra un extracto del resultado correspondiente al dataset “url\_t”.

URL_ID	EXTENDED
1	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
2	<a href="https://twitter.com/WebDesignMagz/status/87963982547259393">https://twitter.com/WebDesignMagz/status/87963982547259393</a>
3	<a href="http://www.d.umn.edu/itss/training/online/webdesign/webdev_listserv.html#subscribe">http://www.d.umn.edu/itss/training/online/webdesign/webdev_listserv.html#subscribe</a>
4	<a href="http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html">http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html</a>
5	<a href="http://www.w3.org/html/wg/wiki/ChangeProposal/meta_name%3Dgenerator_does_not_make_missing_">http://www.w3.org/html/wg/wiki/ChangeProposal/meta_name%3Dgenerator_does_not_make_missing_</a>
6	<a href="https://twitter.com/phrincon/statuses/89273089036267520">https://twitter.com/phrincon/statuses/89273089036267520</a>
7	<a href="http://www.sitepoint.com/real-world-accessibility-html5-aria-and-the-modern-web/">http://www.sitepoint.com/real-world-accessibility-html5-aria-and-the-modern-web/</a>
8	<a href="http://site.pheddo.com/">http://site.pheddo.com/</a>
9	<a href="http://www.w3.org/QA/2011/05/html5-lastcall-accessibility.html">http://www.w3.org/QA/2011/05/html5-lastcall-accessibility.html</a>
10	<a href="http://www.w3.org/WAI/PF/HTML/wiki/Work_Topics">http://www.w3.org/WAI/PF/HTML/wiki/Work_Topics</a>

Figura 27: Urls extendidas a ser analizadas.

<sup>10</sup> Ver anexo 1.2.

- **Urls a procesar**

A continuación, se necesita conocer cuáles de las urls extendidas anteriormente recuperadas se compartieron en dos o más tweets y para generar esa información se realiza una consulta a la base de datos relacionando las tablas url\_t, linked y twett<sup>11</sup>.

Como resultado se obtiene un total de 22409 urls extendidas, y a continuación se presenta un extracto del número total de tweets que compartieron una misma url.

URL_ID	num
1	16
3	157
4	2
7	2
8	2
9	6
10	4
11	2
13	5
16	4
17	2
20	33
22	4
24	2
27	3
29	130
30	2

Figura 28: Tweets que comparten una misma URL

Por ejemplo se presenta tres URL\_ID con su respectiva url extendida:

URL\_ID:1=<http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/> se compartió en 16 tweets.

URL\_ID:3=[http://www.d.umn.edu/itss/training/online/webdesign/webdev\\_listserv.html#subscribe](http://www.d.umn.edu/itss/training/online/webdesign/webdev_listserv.html#subscribe) se compartió en 153 tweets.

URL\_ID:4=<http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html> se compartió en 2 tweets.

---

<sup>11</sup> Ver anexo 3.2.

### 2.3.3.3. Recuperar tweets

Al conocer las urls extendidas que deben ser procesadas, se recupera todos los tweets que compartieron una misma url. La tabla “twett” en su atributo TEXT tiene almacenado el tweet, se necesita relacionar las tablas url\_t, linked, y tweet para obtener todos los tweets<sup>12</sup>.

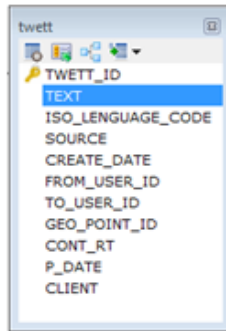


Figura 29: Tabla tweet

En este caso se debe recuperar 16 tweets que compartieron la URL\_ID=1, 157 tweets que compartieron la URL\_ID=2, 2 tweets que compartieron la URL\_ID=4, y así sucesivamente hasta terminar de obtener todos los tweets que comparten una misma url.

Como ejemplo se indica el resultado de los 16 tweets, que comparten la “URL\_ID=1”, como se lo indica a continuación:

URL_ID	TWETT_ID	TEXT
1	8850	#HTML5 HTML5 Accessibility Chops: form control labeling   The Paciello ... What the HTML5 spec... http://bit.ly/izN541 #HTM
1	8815	RT @webdesignmagz: HTML5 Accessibility Chops: form control labeling   The Paciello ... http://bit.ly/kn6QRR #web... http://b
1	8820	RT @WebDesignMagz: HTML5 Accessibility Chops: form control labeling   The Paciello ... http://bit.ly/kn6QRR #webdesign
1	8894	HTML5 Accessibility Chops: form control labeling   The Paciello ... http://bit.ly/kn6QRR #webdesign via:@WebDesignMagz
1	8899	HTML5 Accessibility Chops: form control labeling   The Paciello ... http://bit.ly/kn6QRR #webdesign
1	8891	#clip The Paciello Group Blog - HTML5 Accessibility Chops: form control labeling http://bit.ly/m8ofT1
1	8557	RT @stevefaulkner: HTML5 Accessibility Chops: form control labeling http://t.co/89qrrW1 /cc @davidbolter @jcasteh
1	8927	? @webaxe: HTML5 Accessibility Chops: form control labeling by @stevefaulkner - http://t.co/89qrrW1 #ally #forms #html5
1	9221	RT @webaxe: HTML5 Accessibility Chops: form control labeling by @stevefaulkner - http://t.co/89qrrW1 #ally #forms #html5
1	9452	RT @webaxe: HTML5 Accessibility Chops: form control labeling by @stevefaulkner - http://t.co/89qrrW1 #ally #forms #html5
1	9711	HTML5 Accessibility Chops: form control labeling by @stevefaulkner - http://t.co/89qrrW1 #ally #forms #html5
1	9743	RT @stevefaulkner #HTML5 Accessibility Chops: form control labeling http://t.co/89qrrW1 /cc @davidbolter @jcasteh
1	9744	RT @stevefaulkner #HTML5 Accessibility Chops: form control labeling http://t.co/89qrrW1 /cc @davidbolter @jcasteh
1	9773	HTML5 Accessibility Chops: form control labeling http://t.co/89qrrW1 /cc @davidbolter @jcasteh
1	19473041	Useful read particularly the screen reader tests - form control labeling by @stevefaulkner http://t.co/6j0BBYHS #accessibili
1	20157558	Form control labelling and accessibility by @stevefaulkner http://t.co/xYpliXSH #formsdev via @RuthEllison

Figura 30: Tweets que comparten la URL\_ID=1

Asimismo se presenta parte de los resultados de los tweets que compartieron “URL\_ID=3”, donde se obtiene un registro de 153 tweets.

<sup>12</sup> Ver anexo 3.3.

URL_ID	TWETT_ID	TEXT
3	17563	RT @laura_carlson: Today's Web Design Update http://url.ca/4nx2l Subscribe info: http://url.ca/178he #Accessibility #CSS #HTML5 #usability
3	17585	Today's Web Design Update http://url.ca/4nx2l Subscribe info: http://url.ca/178he #Accessibility #CSS #HTML5 #usability
3	333360	Today's Web Design Update http://t.co/SS51VVRT Subscribe info: http://t.co/oSrVkgG1 #accessibility #CSS #HTML5
3	339445	RT @laura_carlson: Today's Web Design Update http://t.co/OIKV9Qoi Subscribe info: http://t.co/AjgB7BHT #accessibility #A1
3	721560	RT @laura_carlson: Today's Web Design Update http://t.co/3oH2ftxR Subscribe info: http://t.co/AjgB7BHT #accessibility #A1
3	715227	Today's Web Design Update http://t.co/m9dPVRQk Subscribe info: http://t.co/oSrVkgG1 #accessibility #CSS #HTML5
3	714687	RT @aardrian: I'm in it! RT @laura_carlson: Today's Web Design Update http://t.co/jnnBbJx4 Subscribe info: http://t.co/NUrA6wbk #access
3	714737	I'm in it! RT @laura_carlson: Today's Web Design Update http://t.co/jnnBbJx4 Subscribe info: http://t.co/NUrA6wbk #access
3	1034549	Today's Web Design Update http://t.co/MAHIJOYJ Subscribe info: http://t.co/oSrVkgG1 #accessibility #CSS #HTML5 #webdesign
3	1215338	RT @laura_carlson: Today's Web Design Update http://t.co/QFuAPoLP Subscribe info: http://t.co/oSrVkgG1 #accessibility #CS
3	2298963	RT @laura_carlson: Today's Web Design Update http://t.co/5VfYmstM Subscribe info: http://t.co/oSrVkgG1 #accessibility #CS
3	2299010	RT @laura_carlson: Today's Web Design Update http://t.co/5VfYmstM Subscribe info: http://t.co/oSrVkgG1 #accessibility #CS
3	2299022	Today's Web Design Update http://t.co/5VfYmstM Subscribe info: http://t.co/oSrVkgG1 #accessibility #CSS #HTML5 #PHP
3	5491102	RT @laura_carlson: Today's Web Design Update http://t.co/KENt6RbD Subscribe info: http://t.co/oSrVkgG1 #accessibility #CS
3	5491104	Today's Web Design Update http://t.co/KENt6RbD Subscribe info: http://t.co/oSrVkgG1 #accessibility #CSS #webdesign

Figura 31: Tweets que comparten la URL\_ID=3

La URL\_ID=4 se compartió en dos tweets, Este proceso se lo realiza hasta recuperar todos los tweets que comparten una misma url, se trabaja específicamente con el TWETT\_ID.

URL_ID	TWETT_ID	TEXT
4	17563	RT @laura_carlson: Today's Web Design Update http://url.ca/4nx2l Subscribe info: http://url.ca/178he #Accessibility #CSS #HTML5 #usability
4	17585	Today's Web Design Update http://url.ca/4nx2l Subscribe info: http://url.ca/178he #Accessibility #CSS #HTML5 #usability

Figura 32: Tweets que comparten la URL\_ID=4

#### 2.3.3.4. Recuperar etiquetas, menciones, urls, texto

Es de nuestro interés acceder a los registros de las etiquetas, menciones, urls y texto que pertenecen a los tweets, y para recuperar esa información se realiza las siguientes operaciones: se accede a la base de datos<sup>13</sup> seleccionando las tablas que almacenan los atributos antes mencionados y se obtiene los datos que correspondan a un TWETT\_ID basado en una URL\_ID, este proceso se lo debe realizar por cada uno de los tweets recuperados, para la demostración se ha utilizado la url extendida "URL\_ID=1".

- **Recuperar etiquetas.**

La tabla "tag\_t" tiene almacenado el atributo NAME que se refiere al nombre de las etiquetas que son parte del tweet.

tag_t
TAG_ID
NAME

Figura 33: Tabla etiquetas

<sup>13</sup> Ver anexo 3.1.

Se recupera cada una de las etiquetas que pertenecen a un mismo tweet, se utiliza las tablas url\_t, linked, twett, tagged, tag\_t. que tienen la información necesaria para obtener todas las etiquetas.

Se presenta el resultado de todas las etiquetas de los 16 tweets que compartieron la URL=1.

URL_ID	TWETT_ID	NAME
1	8850	HTML5
1	8850	HTML
1	8850	webdesign
1	8815	web
1	8820	webdesign
1	8894	webdesign
1	8899	webdesign
1	8891	clip
1	8927	a11y
1	8927	forms
1	8927	HTML5
1	9221	a11y
1	9221	forms
1	9221	HTML5
1	9452	a11y
1	9452	forms
1	9452	HTML5
1	9711	a11y
1	9711	forms
1	9711	HTML5
1	9743	HTML5
1	9744	HTML5
1	19473041	accessibility
1	19473041	forms
1	20157558	formsdev

Figura 34: Etiquetas de los tweets

Para realizar el tratamiento de las etiquetas es necesario concatenarlas<sup>14</sup> por el TWETT\_ID para posteriormente poder realizar una comparación.

Tabla 3: Etiquetas concatenadas por el TWETT\_ID de la URL\_ID=1

URL_ID	TWETT_ID	NAME
1	8850	HTML5, HTML, webdesign
1	8815	Web
1	8820	Webdesign
1	8894	Webdesign

<sup>14</sup> Ver anexo 2.2

1	8899	Webdesign
1	8891	Clip
1	8927	a11y, forms, HTML5
1	9221	a11y, forms, HTML5
1	9452	a11y, forms, HTML5
1	9711	a11y, forms, HTML5
1	9743	HTML5
1	9744	HTML5
1	19473041	accessibility, forms
1	20157558	Formsdev

- **Recuperar menciones.**

En la tabla “user\_t” se encuentra el atributo NAME que tiene almacenado el nombre de las menciones que son parte del tweet.

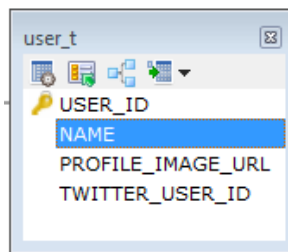


Figura 35: Tabla menciones

Al recuperar las menciones que pertenecen a los tweets que compartieron una misma url, se utiliza las siguientes tablas: url\_t, linked, twett, mention, user\_t. Las menciones de los tweets que comparten la URL\_ID =1, son:

URL_ID	TWEET_ID	NAME
1	8815	webdesignmagz
1	8820	webdesignmagz
1	8894	webdesignmagz
1	8557	davidbolter
1	8557	jcsteh
1	8557	stevefaulkner
1	8927	stevefaulkner
1	8927	webaxe
1	9221	stevefaulkner
1	9221	webaxe

Figura 36: Menciones de los tweets

Para poder comparar se concatena todas las menciones<sup>15</sup> que pertenezcan a un mismo TWETT\_ID como se lo indica en la siguiente tabla.

Tabla 4: Menciones concatenadas por el TWETT\_ID de la URL\_ID=1

URL_ID	TWETT_ID	NAME
1	8815	webdesignmagz
1	8820	webdesignmagz
1	8894	webdesignmagz
1	8557	davidbolter, jcsteh, stevefaulkner
1	8927	stevefaulkner, webaxe
1	9221	stevefaulkner, webaxe
1	9452	stevefaulkner, webaxe
1	9711	stevefaulkner
1	9743	davidbolter, jcsteh, stevefaulkner
1	9744	davidbolter, jcsteh, stevefaulkner
1	9773	davidbolter, jcsteh
1	19473041	stevefaulkner
1	20157558	stevefaulkner, RuthEllison

- **Recuperar urls.**

En la tabla “url\_t” el atributo EXTENDED almacena todas las urls extendidas de un tweet.

url_t
URL_ID
EXTENDED
STATUS_CODE
CONTENT_TYPE
STATUS_MESSAGE
STANDARD
TITLE
DESCRIPTION
LANGUAGE
CHAR_SET

Figura 37: Tabla urls

Para recuperar las urls que corresponden a los tweets que compartieron una misma url, se trabaja con las tablas url\_t, linked, y twett.

<sup>15</sup> Ver anexo 2.1



En la tabla se puede apreciar las urls cortas y urls extendidas de los tweets que compartieron la URL=1. Como se observa existen más de dos urls cortas diferentes pero en sí hablan del mismo tema esto se lo puede comprobar con su correspondiente url extendida. En este caso para el tratamiento de las urls se trabaja con urls extendidas de cada tweet, las mismas que permitirán medir los cambios reales de los tweets.

Tabla 5: Urls cortas y Urls extendidas

TWEET_ID	URL CORTA	URL EXTENDIDA
8850	<a href="http://bit.ly/izN54I">http://bit.ly/izN54I</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8815	<a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8815	<a href="http://bit.ly/mchw6I">http://bit.ly/mchw6I</a>	<a href="https://twitter.com/WebDesignMagz/status/87963982547259393">https://twitter.com/WebDesignMagz/status/87963982547259393</a>
8820	<a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8894	<a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8899	<a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8891	<a href="http://bit.ly/m8ofT1">http://bit.ly/m8ofT1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8557	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8927	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9221	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9452	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9711	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9743	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9744	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9773	<a href="http://t.co/89qrrW1">http://t.co/89qrrW1</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
19473041	<a href="http://t.co/6jOBBYH5">http://t.co/6jOBBYH5</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
20157558	<a href="http://t.co/xYp1iXSH">http://t.co/xYp1iXSH</a>	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>

Para comparar se concatena todas las urls<sup>16</sup> que pertenezcan a un mismo TWEET\_ID como se lo indica a continuación:

Tabla 6: Urls extendidas de los tweets que compartieron la URL\_ID=1

TWEET_ID	URL EXTENDIDA
8850	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8815	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a> , <a href="https://twitter.com/WebDesignMagz/status/87963982547259393">https://twitter.com/WebDesignMagz/status/87963982547259393</a>
8820	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8894	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8899	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8891	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8557	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
8927	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9221	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9452	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9711	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9743	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9744	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
9773	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
19473041	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>
20157558	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-form-control-labeling/</a>

- **Recuperar texto**

En la base de datos no se encuentra almacenado el texto del tweet, por tal razón se trabaja con los tweets que fueron recuperados anteriormente en el apartado 2.5.3.3. Se recupera los tweets<sup>17</sup> que compartieron una misma url.

---

<sup>16</sup> Ver anexo 2.3.

En los tweets recuperados es necesario realizar un proceso que aisle todas las menciones, etiquetas, urls con la finalidad de obtener solamente las palabras del tweet<sup>18</sup>, con las funciones de java `replaceAll()` se reemplaza todas las cadenas que coinciden con la expresión regular dada, `replaceFirst()` reemplaza sólo la primera cadena que coincida con la expresión regular encontrada por la cadena a ser reemplazada, `replace()` sustituye la cadena determinada por el patrón con la cadena de reemplazo dado. Para todos los `replace` en donde encuentre el patrón de las expresiones lo reemplaza con un espacio (" ").

- **Procesamiento de texto**

Dentro del contenido del tweet existe datos que son importantes, las menciones `@usuario` se las utiliza para hablar o referir a un usuario, las `#etiquetas` para marcar palabras clave que cataloguen un tópico específico, y por último los enlaces o vínculos que le dan mayor puntuación debido al contenido completo que abarca sobre el tema.

Se extrae toda esta información en base a la necesidad de recuperar solamente el texto del contenido del tweet, el proceso empieza buscando todas las menciones, etiquetas, urls y signos para sustraerlos y reemplazarlos por un espacio, se utiliza la función `substring` que devuelve parte de una expresión, para extraer una palabra se toma como referencia la posición del índice inicial hasta encontrar la posición del índice final, para determinar la posición inicial se trabaja con la función de java `.indexOf()` de la clase `String` de Java, que permite obtener el índice de comienzo de una subcadena dentro de otra, mediante un número entero. En el caso de que el carácter o la cadena no existan devuelve un `-1`. Entonces lo que se quiere es extraer las sucesivas apariciones de una subcadena, para esto se llama la función repetidamente, pero con un segundo parámetro que incluirá el índice de la última posición encontrada más 1. En cada parámetro se muestra los índices que se utilizaron para sustituir lo antes indicado.

---

<sup>17</sup> Ver anexo 2.4

<sup>18</sup> Ver anexo 10.1

**Extraer menciones.-** La mención se distingue de una cuenta común en twitter por el signo de @, en la expresión se indica que a partir del signo toma como inicio de una mención seguido de caracteres sin espacio, hasta que encuentre dos puntos o un espacio finaliza la mención. Para extraer las menciones<sup>19</sup> el índice inicial es el signo “@” y los índices finales son “:” y “espacio en blanco”.

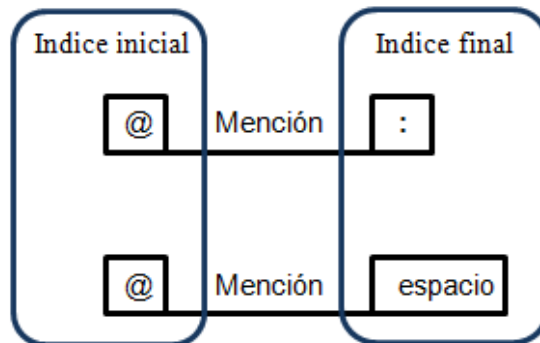


Figura 38: Índices de las menciones

Se presenta un ejemplo con uno de los tweets.

Tabla 7: Extraer menciones

<p>Busca la posición del índice inicial y final.</p>	<p>RT <span style="border: 1px solid black; padding: 2px;">@webdesignmagz:</span> HTML5 Accessibility Chops: form control labeling   The Paciello ... <a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> #web... <a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a></p>
<p>Se sustrae y reemplaza por un espacio en blanco.</p>	<p style="text-align: center;"> <span style="border: 1px solid black; padding: 2px;">@webdesignmagz:</span> → <span style="border: 1px solid black; display: inline-block; width: 40px; height: 20px; vertical-align: middle;"></span> </p>
<p>Resultado</p>	<p>RT HTML5 Accessibility Chops: form control labeling   The Paciello ... <a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> #web... <a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a></p>

<sup>19</sup> Ver anexo 10.3

**Extraer etiquetas.-** Un hashtag o etiqueta se distingue por el signo # la expresión toma este signo como el inicio de la etiqueta seguido de caracteres o números y termina cuando encuentra un espacio. Para extraer las etiquetas<sup>20</sup> se indica que el índice inicial es el signo “#” y el índice final es un “espacio”.

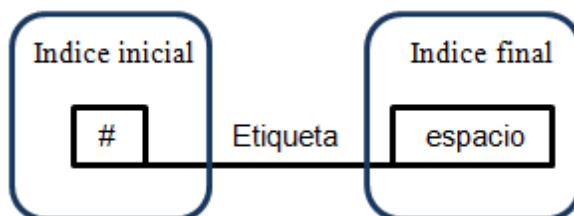


Figura 39: Índices de etiquetas

En el ejemplo se sigue trabajando con el mismo tweet, notando que este recibe el tweet resultante después de haber reemplazado las menciones.

Tabla 8: Extraer etiquetas

<p>Busca la posición del índice inicial y final.</p>	<p>RT HTML5 Accessibility Chops: form control labeling   The Paciello ... <a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> #web... <a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a></p>
<p>Se sustrae y reemplaza por un espacio en blanco.</p>	<p>El diagrama muestra un recuadro azul con el texto '#web'. Una flecha azul apunta desde este recuadro hacia la derecha, donde se encuentra un recuadro azul vacío.</p>
<p>Resultado</p>	<p>RT HTML5 Accessibility Chops: form control labeling   The Paciello ... <a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> ... <a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a></p>

<sup>20</sup> Ver anexo 10.3

**Extraer urls.-** La url se distingue por encontrar http:// o https:// dentro del contenido del tweet, la mayor parte de ellas se encuentran acortadas para reducir el número de caracteres, para la formulación de la expresión toma estos caracteres como el inicio de una url seguido de caracteres, números, signos y termina cuando encuentra un espacio, para extraer las urls<sup>21</sup> los índices iniciales son “http://” y “https://” mientras que el índice final es un “espacio”.

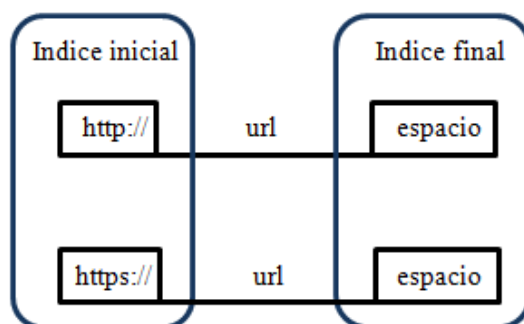


Figura 40: Índices de urls

Como entrada se recibe el tweet resultante después de haber reemplazado las etiquetas.

Tabla 9: Extraer urls

<p>Busca la posición del índice inicial y final.</p>	<p>RT HTML5 Accessibility Chops: form control labeling   The Paciello ... <a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> ... <a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a></p>
<p>Se sustrae y reemplaza por un espacio en blanco.</p>	<p><a href="http://bit.ly/kn6QRR">http://bit.ly/kn6QRR</a> → <input type="text"/></p> <p><a href="http://bit.ly/mchw6l">http://bit.ly/mchw6l</a> → <input type="text"/></p>
<p>Resultado</p>	<p>RT HTML5 Accessibility Chops: form control labeling   The Paciello ... ..</p>

<sup>21</sup> Ver anexo 10.4

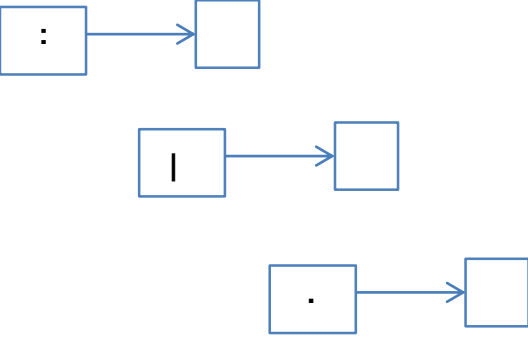
**Extraer signos.-** En los tweets también existen signos a estos se los reemplaza por un espacio<sup>22</sup>, los signos que se han tomado en cuenta son:

Tabla 10: Signos a reemplazar

(	)	?	,	*	&	\$	;	%	^	-
"	'	.	:	/	»	\	!		+	[

De igual forma se recibe el tweet resultante después de haber reemplazado las urls.

Tabla 11: Extraer signos

Busca los signos	RT HTML5 Accessibility Chops: form control labeling   The Paciello ... ..
Reemplaza por un espacio en blanco.	
Resultado	RT HTML5 Accessibility Chops form control labeling The Paciello

<sup>22</sup> Ver anexo 10.2

<sup>23</sup> Ver anexo 10.5

Para la demostración se ha utilizado los tweets que compartieron la “URL\_ID=1”, esto se realiza mientras se haya terminado de procesar el texto de todos los tweets.

El resultado del texto de los tweets que compartieron la URL=1 es:

Tabla 12: Texto de los tweets que compartieron la URL\_ID=1

TWETT_ID	TEXT
8850	HTML5 Accessibility Chops form control labeling The Paciello What the HTML5 spec
8815	RT HTML5 Accessibility Chops form control labeling The Paciello
8820	RT HTML5 Accessibility Chops form control labeling The Paciello
8894	HTML5 Accessibility Chops form control labeling The Paciello via
8899	HTML5 Accessibility Chops form control labeling The Paciello
8891	The Paciello Group Blog HTML5 Accessibility Chops form control labeling
8557	RT HTML5 Accessibility Chops form control labeling cc
8927	HTML5 Accessibility Chops form control labeling by
9221	RT HTML5 Accessibility Chops form control labeling by
9452	RT HTML5 Accessibility Chops form control labeling by
9711	HTML5 Accessibility Chops form control labeling by
9743	RT Accessibility Chops form control labeling cc
9744	RT Accessibility Chops form control labeling cc
9773	HTML5 Accessibility Chops form control labeling cc
19473041	Useful read particularly the screen reader tests form control labeling by
20157558	Form control labelling and accessibility by via

Al existir palabras como por ejemplo: libro y libros o reloj y relojes son diferentes porque la primer palabra se encuentra en singular mientras que la segunda en plural, las palabras que son singulares al formar su plural añaden generalmente diferentes sufijos –s, -es, haciendo que cambie el número de objetos al que se refieren.

Libro = un libro.

Libros = varios libros.

Existen palabras que sólo se emplean en singular, careciendo de un plural y viceversa palabras que sólo emplean plural careciendo de un singular. Son pocas las palabras que permanecen invariables. Más bien son muchas las palabras que pueden ser singulares y se pueden convertir en plurales.



#### 2.3.4. Determinar cambios.

Para medir los cambios en las métricas acordadas, se obtiene todas las etiquetas<sup>24</sup>, menciones<sup>25</sup>, urls<sup>26</sup> y texto<sup>27</sup> de los tweets que compartieron una misma url, al tener esos datos recuperados se los va añadiendo en cada uno de los arraylist en donde se empieza a comparar cada objeto del array contando el número de repeticiones que tienen los objetos y luego en un array auxiliar se va agregando solo una repetición por objeto para poder contar el número de cambios encontrados.

El valor que retorna es el número total de cambios en cada métrica, además se genera un promedio de todas las métricas. A continuación se presenta un ejemplo de cómo se mide los cambios en las etiquetas.

Ejemplo: Luego de recuperar las etiquetas de un tweet se las almacena en un arraylist para poder realizar las comparaciones.

Tabla 13: ArrayList etiquetas

TWEET_ID	ETIQUETAS
8850	HTML5, HTML, webdesign
8815	web
8820	webdesign
8894	webdesign
8899	webdesign
8891	clip
8927	a11y, forms, HTML5
9221	a11y, forms, HTML5
9452	a11y, forms, HTML5
9711	a11y, forms, HTML5
9743	HTML5
9744	HTML5
19473041	accessibility, forms
20157558	formsdev

---

<sup>24</sup> Ver anexo 3.4

<sup>25</sup> Ver anexo 3.5

<sup>26</sup> Ver anexo 3.6

<sup>27</sup> Ver anexo 3.7

Para determinar el número de similitudes se implementó la clase Tipos<sup>28</sup> que realiza el proceso de contar las repeticiones de las etiquetas<sup>29</sup>, menciones<sup>30</sup>, urls<sup>31</sup> y texto<sup>32</sup> que se encuentran en los arrays de acuerdo a cada métrica.

Tabla 14: Número de repeticiones de etiquetas

NÚMERO DE REPETICIONES	ETIQUETAS
1	HTML5, HTML, webdesign
1	web
3	webdesign
	webdesign
	webdesign
1	clip
4	a11y, forms, HTML5
	a11y, forms, HTML5
	a11y, forms, HTML5
	a11y, forms, HTML5
2	HTML5
	HTML5
1	accessibility, forms
1	formsdev

A continuación en un array auxiliar se deja una sola repetición<sup>33</sup> por cada objeto diferente en cada métrica.

Tabla 15: ArrayList de etiquetas

ETIQUETAS	Número de Repeticiones
HTML5, HTML, webdesign	1
web	1
webdesign	1
clip	1
a11y, forms, HTML5	1
HTML5	1
accessibility, forms	1
formsdev	1

<sup>28</sup> Ver anexo 5

<sup>29</sup> Ver anexo 5.1

<sup>30</sup> Ver anexo 5.2

<sup>31</sup> Ver anexo 5.3

<sup>32</sup> Ver anexo 5.4

<sup>33</sup> Ver anexo 11.1

Se empieza a comparar todos los objetos del array para contar el número de cambios<sup>34</sup> encontrados. Como resultado se obtiene un total de 7 cambios en las etiquetas

Tabla 16: Número de cambios

ETIQUETAS	Número de Cambios
HTML5, HTML, webdesign	7
web	
webdesign	
clip	
a11y, forms, HTML5	
HTML5	
accessibility, forms	
formsdev	

En la siguiente tabla se tiene el resultado de la URL=1.

Tabla 17: Total de cambios en etiquetas, menciones, urls y texto de la URL\_ID=1

URL	Cambios_etiqueta	Cambios_mencion	Cambios_url	Cambios_texto
1	7	6	1	13

Además se calcula el promedio<sup>35</sup> de todos los cambios de cada métrica de acuerdo a la url analizada.

Tabla 18: Promedio de los cambios

PROMEDIO DE LOS CAMBIOS
<p><b>Cambios_etiqueta + Cambios_mencion + Cambios_url+ Cambios_texto</b></p> <p><b>7 + 6 + 1 + 13 = 27</b></p> <p><b>37 / 4 = 6,75</b></p>

<sup>34</sup> Ver anexo 11.2

<sup>35</sup> Ver anexo 11.3

El proceso para determinar los cambios<sup>36</sup>, indicado anteriormente se lo realiza para todas las 22409 urls que se compartieron en dos o más tweets con la finalidad de obtener el número total de cambios por cada una de las métricas.

### 2.3.5. Almacenar resultados

Al obtener el número total de cambios de cada uno de los tweets que comparten una misma url, se crea una nueva tabla en la base de datos SNA\_TWITTER denominada “gp\_resultados”, la misma que servirá para guardar<sup>37</sup> el número de cambios de las etiquetas, menciones, url, texto y el promedio calculado.

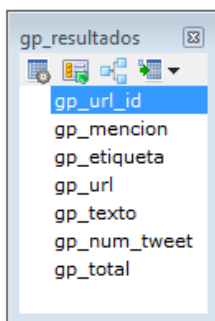


Figura 41: Tabla gp\_resultados

Los datos se almacenarán<sup>38</sup> de la siguiente manera:

- gp\_url\_id: Se refiere al id de la url extendida.
- gp\_mencion: Se refiere al número total de cambios en menciones.
- gp\_etiqueta: Se refiere al número total de cambios en etiquetas.
- gp\_url: Se refiere al número total de cambios en urls.
- gp\_texto: Se refiere al número total de cambios en el texto.
- gp\_num\_tweet: Es el número de tweets que compartieron una misma url extendida.
- gp\_total: Es el promedio general de todos los cambios.

Tabla 19: Resultados de los cambios de la URL\_ID=1

gp_url_id	gp_num_tweet	gp_mencion	gp_etiqueta	gp_url	gp_texto	gp_total
1	16	6	7	1	13	6

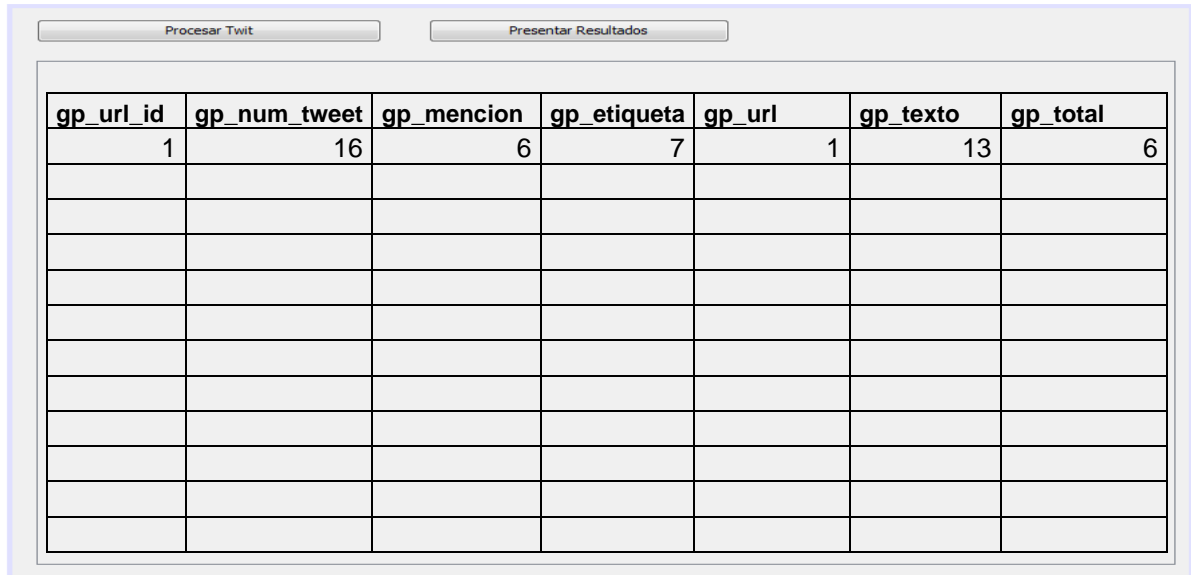
<sup>36</sup> Ver anexo 11

<sup>37</sup> Ver anexo 3.8 y anexo 2.5

<sup>38</sup> Ver anexo 3.9

### 2.3.6. Presentación de resultados

Al terminar de cumplir todas las fases anteriores se deben presentar los resultados<sup>39</sup> en la aplicación, y de la misma manera que se los almacenó en la base de datos se presentarán al usuario, se implementó una tabla<sup>40</sup> en la interfaz de usuario porque permita visualizar de manera clara todos los cambios encontrados durante el proceso del algoritmo.



gp_url_id	gp_num_tweet	gp_mencion	gp_etiqueta	gp_url	gp_texto	gp_total
1	16	6	7	1	13	6

Figura 42: Interfaz para presentar los resultados

<sup>39</sup> Ver anexo 3.10

<sup>40</sup> Ver anexo 12

## **CAPÍTULO 3**

### **ESTUDIO REAL DE LOS CAMBIOS**

En este capítulo se presenta un Estudio real de los cambios, el mismo que detalla un análisis de los resultados que se obtuvieron al ejecutar el algoritmo, se podrá visualizar gráficas estadísticas de los cambios encontrados en los tweets que compartieron una misma url.

Se conocerá cuales urls obtuvieron el mayor y menor número de cambios en cuanto a sus métricas de: menciones, etiquetas, urls, texto. También se conoce cuál es la medida que cambió con mayor frecuencia tomando como referencia su promedio. Una de las combinaciones que se hace es calcular algunas medidas de tendencia central, que permiten identificar los valores más representativos de los datos, de acuerdo a la manera como se tienden a concentrar.

Por ejemplo la Media indica el promedio de los datos; es decir, informa el valor que obtendría cada uno de los individuos si se distribuyeran los valores en partes iguales. La Moda indica el valor que más se repite dentro de los datos. Las medidas de dispersión, también llamadas medidas de variabilidad, muestran la variabilidad de una distribución, indicando por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, cuanto menor sea, más homogénea será a la media. Así se conocerá si todos los casos son parecidos o varían mucho entre ellos.

Como se puede observar en la siguiente figura se presenta un extracto de los resultados almacenados en la base de datos, sin olvidar que los datos representan los cambios totales de cada una de las métricas. Se obtuvo un total de 22409 resultados.

gp_url_id	gp_num_tweet	gp_mencion	gp_etiqueta	gp_url	gp_texto	gp_total
1	16	6	7	1	13	6
3	157	4	21	64	25	28
4	2	0	0	0	1	0
7	2	0	0	0	1	0
8	2	0	0	1	1	0
9	6	2	2	1	4	2
10	4	1	1	0	3	1
11	2	0	1	0	1	0
13	5	0	0	2	3	1
16	4	2	1	0	2	1
17	2	0	0	0	1	0
20	33	12	7	0	23	10
22	4	0	0	0	2	0
24	2	0	0	0	1	0
27	3	0	0	0	1	0
29	130	51	4	1	76	33
30	2	0	0	0	1	0

Figura 43: Resultados de los cambios

Se calcula una de las medidas de dispersión con todo los resultados del promedio de los cambios, denominada Rango que es la diferencia entre el mayor y el menor de los datos de una distribución estadística, su fórmula es la siguiente:

$$Rango = (Max - Min)$$

Figura 44: Fórmula de rango

Tabla 20: Cálculo del rango

MAXIMO	MINIMO	RANGO
140	0	140

Otra medida es Medio rango.

$$medioRango = \frac{(Max + Min)}{2}$$

Figura 45: Fórmula de medio rango

Tabla 21: Cálculo de medio rango

MAXIMO	MINIMO	RANGO
140	0	70

La desviación respecto a la media es la diferencia entre cada valor de la variable estadística y la media aritmética. Para calcular la desviación media, primero se calcula la media aritmética.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3 + 8 + 8 + 8 + 9 + 9 + 9 + 18}{8} = 9$$

Figura 46: Fórmula de la media aritmética.

A continuación se calcula la desviación promedio, con el siguiente ejemplo:

$$DM = \frac{\sum |x - \bar{x}|}{n}$$

$$DM = \frac{|3 - 9| + |8 - 9| + |8 - 9| + |8 - 9| + |9 - 9| + |9 - 9| + |9 - 9| + |18 - 9|}{8}$$

$$DM = \frac{6 + 1 + 1 + 1 + 0 + 0 + 0 + 9}{8} = \frac{18}{8} = 2,25$$

Figura 47: Fórmula de la desviación promedio.

En el promedio de cambios el resultado de la desviación promedio tiene un valor de 1,24412395.



### 3.1. Cambios de las urls que compartieron el mayor número de tweets.

Se analiza las 20 Urls que han compartido el mayor número de tweets, se toma como referencia el atributo "NumTweets".

Tabla 22: Cambios de las 20 urls que se compartieron en más tweets

URL	NumTweets	Cambios Etiquetas	Cambios Menciones	Cambios Urls	Cambios Texto	Promedio Cambios
3365	1525	46	8	2	146	50
27107	1086	4	0	0	12	4
2416	907	156	35	1	328	130
28920	585	157	48	5	250	115
2582	552	395	8	2	52	114
28669	482	154	45	11	259	117
276	436	288	33	1	85	101
180	417	407	1	2	14	106
15053	376	103	37	3	162	76
17236	370	14	49	7	93	40
36882	369	360	0	1	200	140
6743	354	40	36	1	126	50
57304	335	36	27	6	131	50
23713	333	14	48	3	70	33
47678	321	17	35	9	96	39
14156	316	53	37	4	155	62
10183	308	72	38	1	122	58
37205	304	23	16	1	139	44
14341	293	16	37	6	73	33
13881	282	41	35	6	102	46

Las 10 urls que compartieron el mayor número de tweets se encuentran ordenadas por el atributo NumTweets y son las siguientes:

Tabla 23: Urls que compartieron el mayor número de tweets

NumTweets	URL_ID	URL
1525	3365	<a href="https://www.facebook.com/">https://www.facebook.com/</a>
1086	27107	<a href="http://www.imsorrythatwasalie.com/">http://www.imsorrythatwasalie.com/</a>
907	2416	<a href="http://www.petapixel.com/2011/11/29/cnn-lays-off-photojournalists-citing-the-accessibility-of-quality-cameras/">http://www.petapixel.com/2011/11/29/cnn-lays-off-photojournalists-citing-the-accessibility-of-quality-cameras/</a>
585	28920	<a href="http://www.mysqltalk.com/participate.html">http://www.mysqltalk.com/participate.html</a>
552	2582	<a href="http://paper.li/barrierfreemb/1321810432">http://paper.li/barrierfreemb/1321810432</a>
482	28669	<a href="http://www.mysqltalk.com/gaad.html">http://www.mysqltalk.com/gaad.html</a>
436	276	<a href="http://paper.li/Hear_And_There/1306429577">http://paper.li/Hear_And_There/1306429577</a>
417	180	<a href="http://paper.li/umegaki/1313807124">http://paper.li/umegaki/1313807124</a>
376	15053	<a href="https://www.facebook.com/globalaccessibilityawarenessday">https://www.facebook.com/globalaccessibilityawarenessday</a>
370	17236	<a href="http://www.engadget.com/2012/03/01/sprint-and-code-factory-team-up-to-deliver-free-android-accessib/">http://www.engadget.com/2012/03/01/sprint-and-code-factory-team-up-to-deliver-free-android-accessib/</a>

En la figura se indica el resultado de los cambios en menciones, etiquetas, urls, y texto de las urls que han compartido el mayor número de tweets.

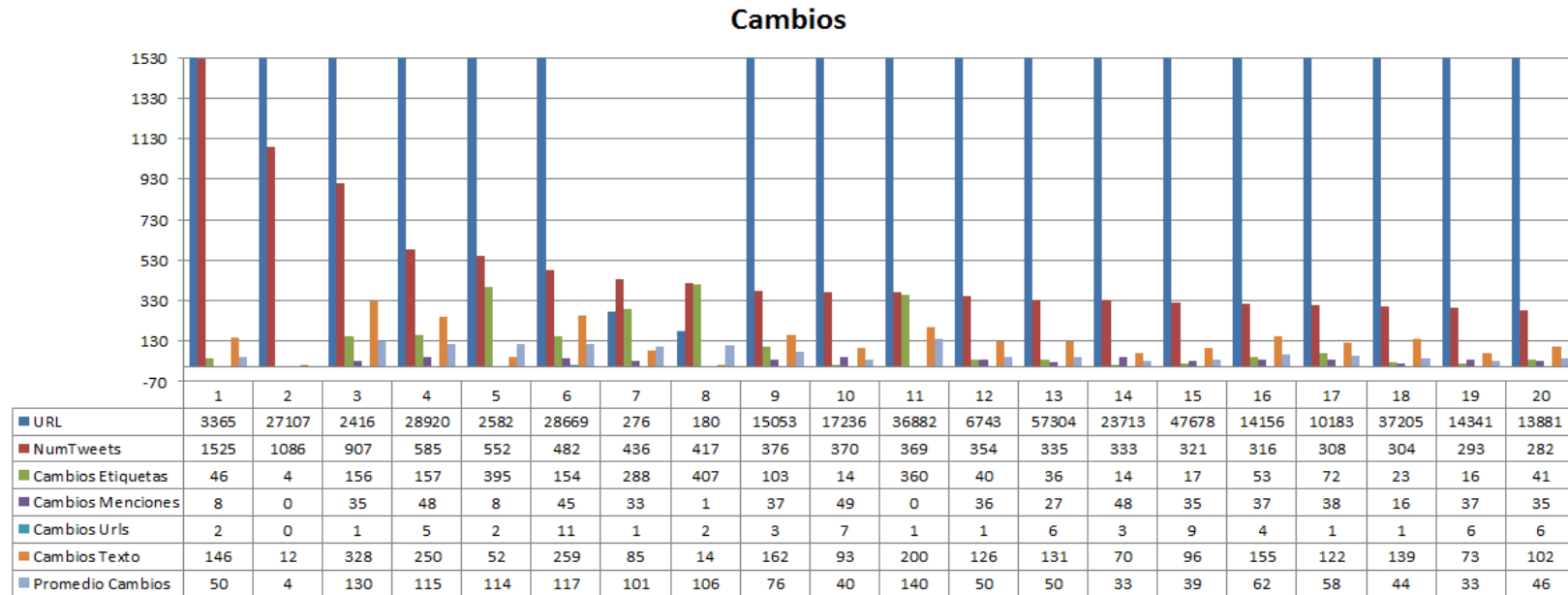


Figura 48: Cambios de las urls que han compartido el mayor número de tweets

La URL 3365= <https://www.facebook.com/> compartió 1525 tweets, y en la siguiente tabla se indica los resultados de los cambios de cada una de las métricas.

Tabla 24: Cambios de la URL\_ID =3365

URL	NumTweets	Cambios Etiquetas	Cambios Menciones	Cambios Urls	Cambios Texto	Promedio Cambios
3365	1525	46	8	2	146	50

Porcentaje total del número de cambios de la url que compartió el mayor número de tweets.

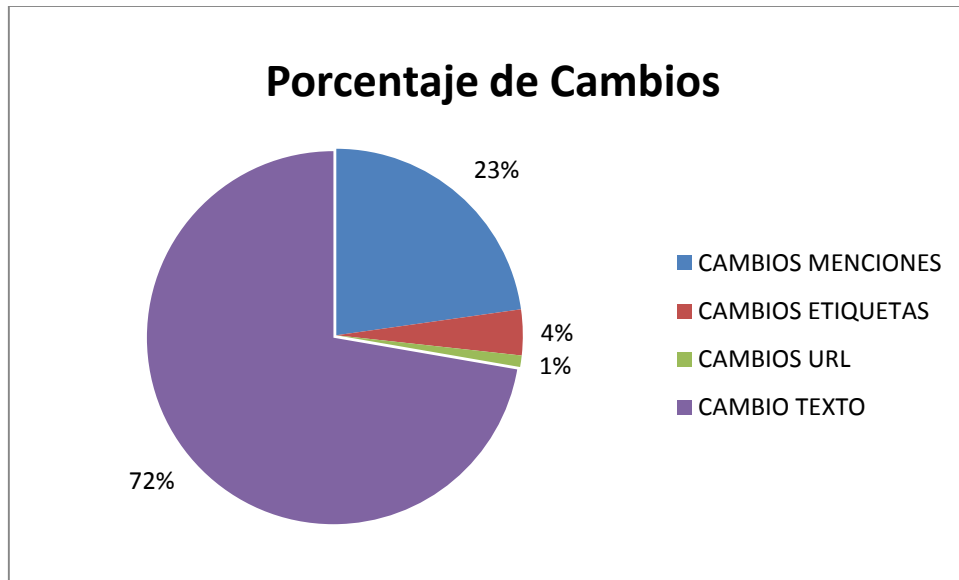


Figura 49: Porcentaje de cambios de la URL=3365

Se obtiene que de un 100%, la métrica que mayor cambio tuvo es el texto, y en la que menor cambio se reflejó es en la url.

Tabla 25: Porcentaje de cambios de la URL\_ID=3365

URL	METRICA	PORCENTAJE
3365 <a href="https://www.facebook.com/">https://www.facebook.com/</a>	Texto	72
	Menciones	23
	Etiqueta	4
	Url	1
	<b>Total</b>	<b>100</b>

### 3.2. Cambios de las urls que compartieron el menor número de tweets.

Al analizar los cambios de las 20 Urls que compartieron el menor número de tweets, se toma como referencia el atributo NumTweets que se encuentra ordenado conjuntamente con la URL.

Tabla 26: Cambios de las 20 urls que compartieron el menor número de tweets

URL	NumTweets	Cambios Etiquetas	Cambios Menciones	Cambios Urls	Cambios Texto	Promedio Cambios
4	2	0	0	0	1	0
7	2	0	0	0	1	0
8	2	0	0	1	1	0
11	2	0	1	0	1	0
17	2	0	0	0	1	0
24	2	0	0	0	1	0
30	2	0	0	0	1	0
37	2	0	0	0	1	0
40	2	0	0	0	0	0
47	2	0	1	0	1	0
61	2	0	1	0	1	0
62	2	0	0	0	0	0
70	2	0	0	0	0	0
79	2	0	0	0	0	0
81	2	1	0	0	1	0
82	2	0	1	0	1	0
83	2	0	1	0	1	0
86	2	0	0	0	0	0
89	2	0	1	0	1	0
93	2	0	0	0	1	0

Las 10 urls que compartieron el menor número de tweets están ordenadas por NumTweets y la URL\_ID.

Tabla 27: Urls que compartieron el menor número de tweets

NumTweets	URL_ID	URL
2	4	<a href="http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html">http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html</a>
2	7	<a href="http://www.sitepoint.com/real-world-accessibility-html5-aria-and-the-modern-web/">http://www.sitepoint.com/real-world-accessibility-html5-aria-and-the-modern-web/</a>
2	8	<a href="http://site.pheedo.com/">http://site.pheedo.com/</a>
2	11	<a href="http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-aria-landmark-support/">http://www.paciellogroup.com/blog/2011/07/html5-accessibility-chops-aria-landmark-support/</a>
2	17	<a href="http://www.w3.org/TR/html5-author/the-img-element.html">http://www.w3.org/TR/html5-author/the-img-element.html</a>
2	24	<a href="http://techfreedom.org/publications/fcc-requires-online-public-inspection-files-misses-point-opengov-data-accessibility">http://techfreedom.org/publications/fcc-requires-online-public-inspection-files-misses-point-opengov-data-accessibility</a>
2	30	<a href="http://webdesign.tutsplus.com/articles/general/accessibility-whats-your-markup-saying-to-you/?t=1320119962">http://webdesign.tutsplus.com/articles/general/accessibility-whats-your-markup-saying-to-you/?t=1320119962</a>
2	37	<a href="http://setaris.com/2011/10/ux-seo-and-accessibility/?t=1320157446">http://setaris.com/2011/10/ux-seo-and-accessibility/?t=1320157446</a>
2	40	<a href="http://owl.li/5yeZl">http://owl.li/5yeZl</a>
2	47	<a href="http://lists.d.umn.edu/pipermail/webdev/2011q4/000616.html">http://lists.d.umn.edu/pipermail/webdev/2011q4/000616.html</a>

En la figura se indica el resultado de todos los cambios en menciones, etiquetas, urls y texto de las urls que han compartido el menor número de tweets.

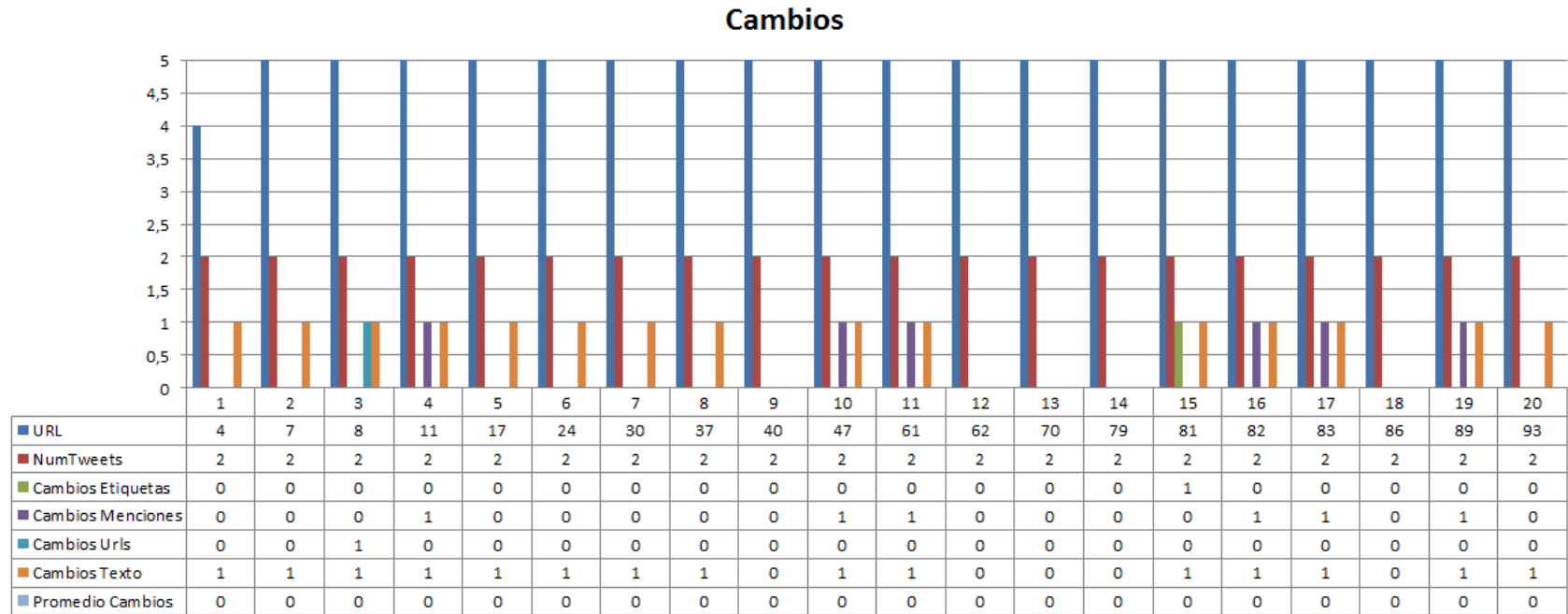


Figura 50: Cambios de las urls que han compartido el menor número de tweets.

La URL 4 = <http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html> compartió 2 tweets, y en la siguiente tabla se indica los resultados de los cambios de cada una de las métricas.

Tabla 28: Cambios de la URL\_ID =4

URL	NumTweets	Cambios Etiquetas	Cambios Menciones	Cambios Urls	Cambios Texto	Promedio Cambios
4	2	0	0	0	1	0

### 3.3. Urls con mayor y menor número de cambios en menciones.

Al realizar un estudio más específico, se empieza analizando el número de cambios que presenta las menciones de los tweets que compartieron una misma url. Mostrando un resultado con 25 datos ordenados por “Cambios Menciones” para determinar el mayor y menor número de cambios en esta métrica, como se lo indica en la siguiente tabla.

Tabla 29: Urls con mayor y menor número de cambios en menciones

	URL	Cambios Menciones		URL	Cambios Menciones
	Url con mayor número de cambios en Menciones	180		407	Url con menor número de cambios en Menciones
2582		395	7	0	
36882		360	8	0	
276		288	11	0	
13377		195	13	0	
28920		157	17	0	
2416		156	22	0	
28669		154	24	0	
1810		135	27	0	
8591		111	30	0	
15053		103	37	0	
10183		72	40	0	
1941		70	44	0	
334		69	47	0	
31		65	58	0	
3742		62	60	0	
44936		60	61	0	
293		58	62	0	
34365		54	70	0	
597		53	79	0	
14156		53	82	0	
29		51	83	0	
1614		51	86	0	
4853		47	88	0	
3365		46	89	0	

En la figura se muestra el resultado del mayor número de cambios en menciones que han compartido una misma url extendida.

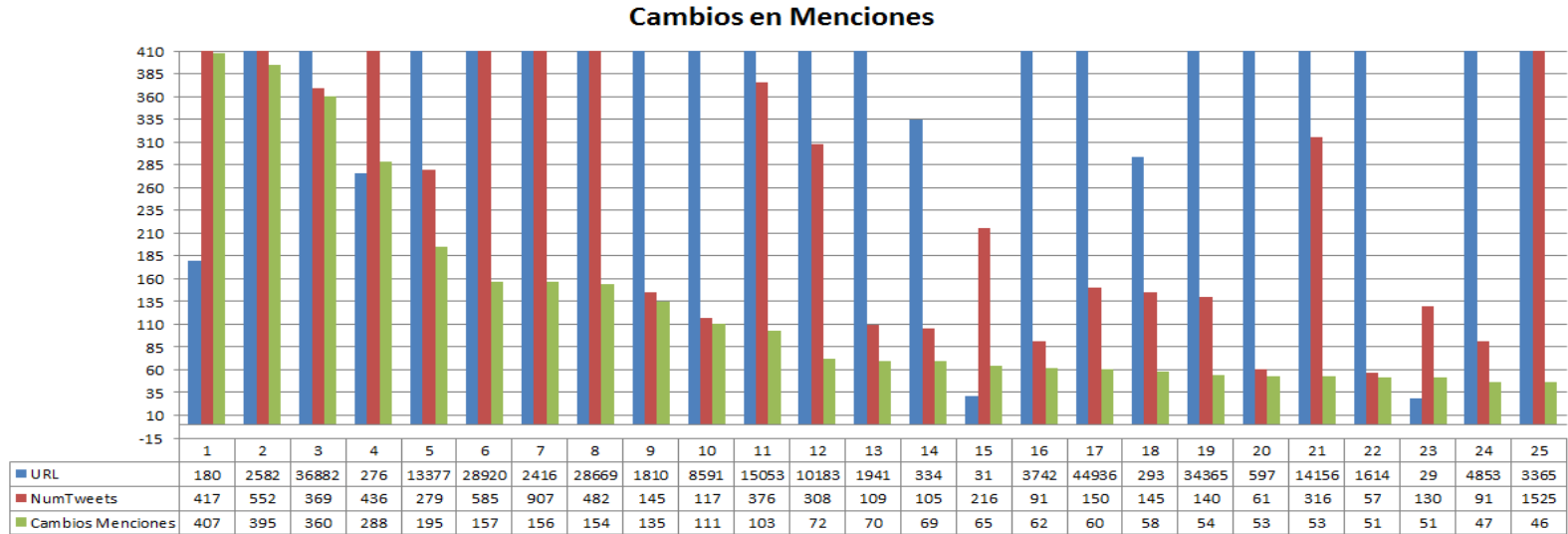


Figura 51: Urls con el mayor número de cambios en menciones

Se indica el resultado del menor número de cambios en menciones que han compartido una misma url extendida.

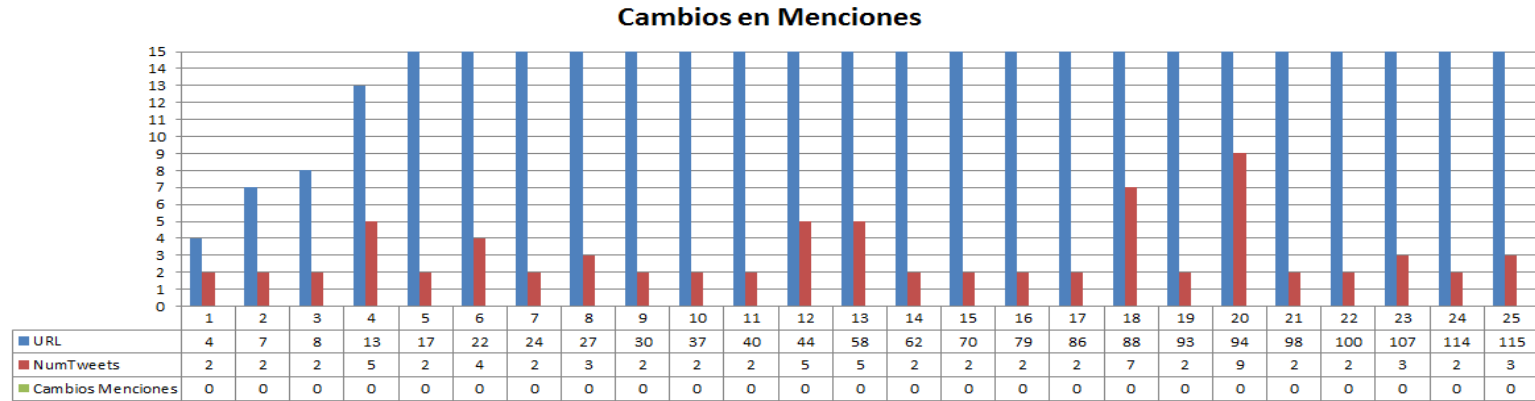


Figura 52: Urls con el menor número de cambios en menciones

Los tweets que compartieron la URL 180= <http://paper.li/umegaki/1313807124>, fueron los que mayor cambio tuvieron con respecto a las menciones en donde existió 407 cambios de un total de 417 tweets, y la URL 4 = <http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html> que fue compartida en dos tweets no cambió las menciones.

A continuación se indica la información de la url con mayor y menor número de cambios en menciones.

Tabla 30: Url con mayor número de cambios en mención

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
180	417	407	1	2	14	106

Porcentaje del total de cambios del tweet con mayor cambio en menciones.

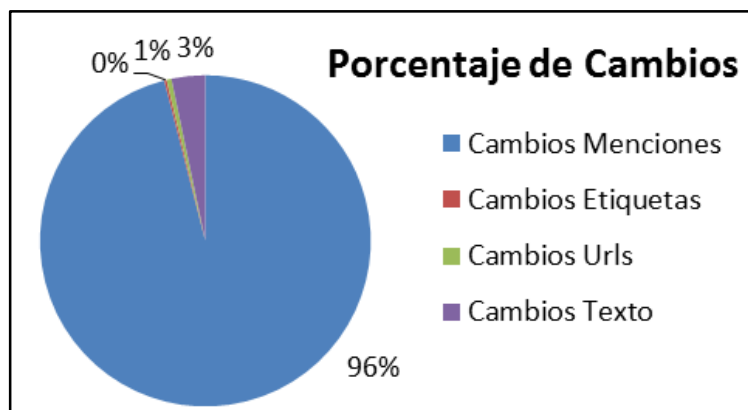


Figura 53: Porcentaje de cambios de la URL=180

Tabla 31: Url con menor número de cambios en mención

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
4	2	0	0	0	1	0

En la tabla anterior se observa que no hubo cambio en menciones, etiquetas y urls pero si existió un cambio en el texto.



### 3.4. Urls con mayor y menor número de cambios en etiquetas.

Continuando con el estudio, se analiza el número de cambios que existe en las etiquetas de los tweets que compartieron una misma url. Se muestra un resultado con 25 datos ordenados por “Cambios Etiquetas” para determinar el mayor y menor número de cambios de esta métrica, como se lo indica en la siguiente tabla.

Tabla 32: Urls con mayor y menor número de cambios en etiquetas

Url con mayor número de cambios en Etiquetas	URL	Cambios Etiquetas	Url con menor número de cambios en Etiquetas	URL	Cambios Etiquetas
	17236	49		4	0
	23713	48		7	0
	28920	48		8	0
	28669	45		13	0
	31	43		17	0
	16866	39		22	0
	10183	38		24	0
	41340	38		27	0
	14156	37		30	0
	14341	37		34	0
	15053	37		37	0
	6743	36		38	0
	2416	35		40	0
	13881	35		41	0
	47678	35		44	0
	276	33		52	0
	40607	33		58	0
	2493	31		62	0
	49017	31		70	0
	49186	29		79	0
	526	28		81	0
	47899	28		86	0
	57304	27		88	0
	2397	26		93	0
22447	26	94	0		

En la figura se muestra el resultado del mayor número de cambios en etiquetas que han compartido una misma url extendida.

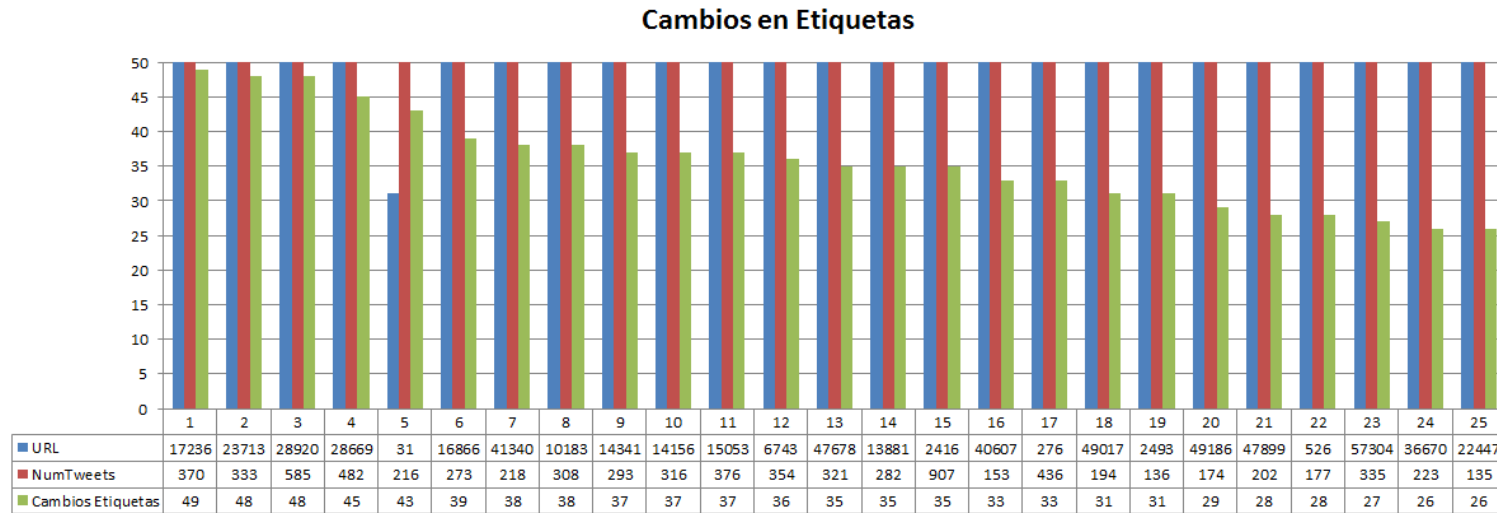


Figura 54: Urls con el mayor número de cambios en etiquetas

En la figura se observa el resultado del menor número de cambios en etiquetas que han compartido una misma url extendida.

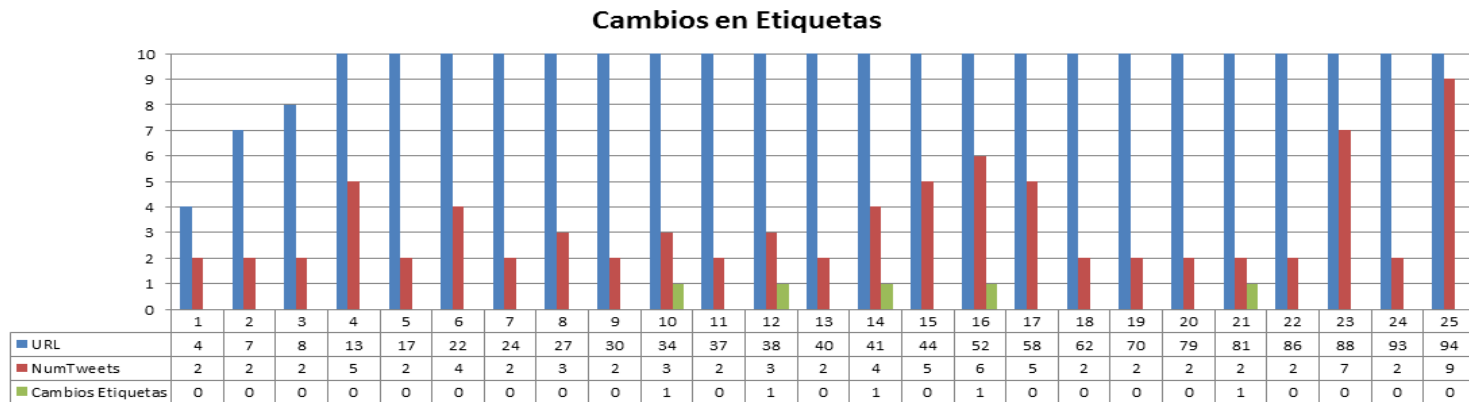


Figura 55: Urls con el menor número de cambios en etiquetas

Los tweets que compartieron la URL 17236 = <http://www.engadget.com/2012/03/01/sprint-and-code-factory-team-up-to-deliver-free-android-accessib/>, fueron los que mayor cambio tuvieron en las etiquetas en donde existió 49 cambios de un total de 370 tweets, y la URL 4 = <http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html> que fue compartida en dos tweets no cambió las etiquetas.

A continuación se indica la información de la url con mayor y menor número de cambios en las etiquetas.

Tabla 33: Url con mayor número de cambios en etiquetas

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
17236	370	14	49	7	93	40

Porcentaje del total de cambios del tweet con mayor cambio en etiquetas.

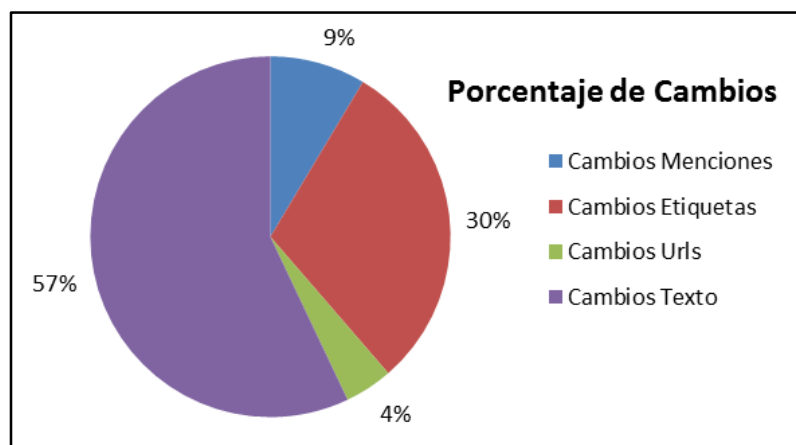


Figura 56: Porcentaje de cambios de la URL=17236

Tabla 34: Url con menor número de cambios en etiquetas

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
4	2	0	0	0	1	0

### 3.5. Urls con mayor y menor número de cambios en urls

Se analiza el número de cambios que existe en las urls que contienen los tweets que compartieron una misma url. Se indica un resultado con 25 datos ordenados por “Cambios Urls” para determinar el mayor y menor número de cambios en esta métrica, como se lo indica en la siguiente tabla.

Tabla 35: Urls con mayor y menor número de cambios en urls

Url con mayor número de cambios en Urls	URL	Cambios Urls	Url con menor número de cambios en Urls	URL	Cambios Urls
	3	64		4	0
	27876	59		7	0
	15664	53		8	0
	9805	46		13	0
	5213	46		17	0
	20798	25		22	0
	7225	21		24	0
	1056	17		27	0
	42463	16		30	0
	249	14		34	0
	22678	13		37	0
	293	13		38	0
	25642	12		40	0
	1906	11		41	0
	28669	11		44	0
	29015	9		52	0
	46278	9		58	0
	41214	9		62	0
	47678	9		70	0
	5994	8		79	0
	6000	8		81	0
	45016	8		86	0
	47281	8		88	0
	1471	8		93	0
9319	7	94	0		

En la figura se demuestra el resultado del mayor número de cambios en urls que han compartido una misma url extendida.

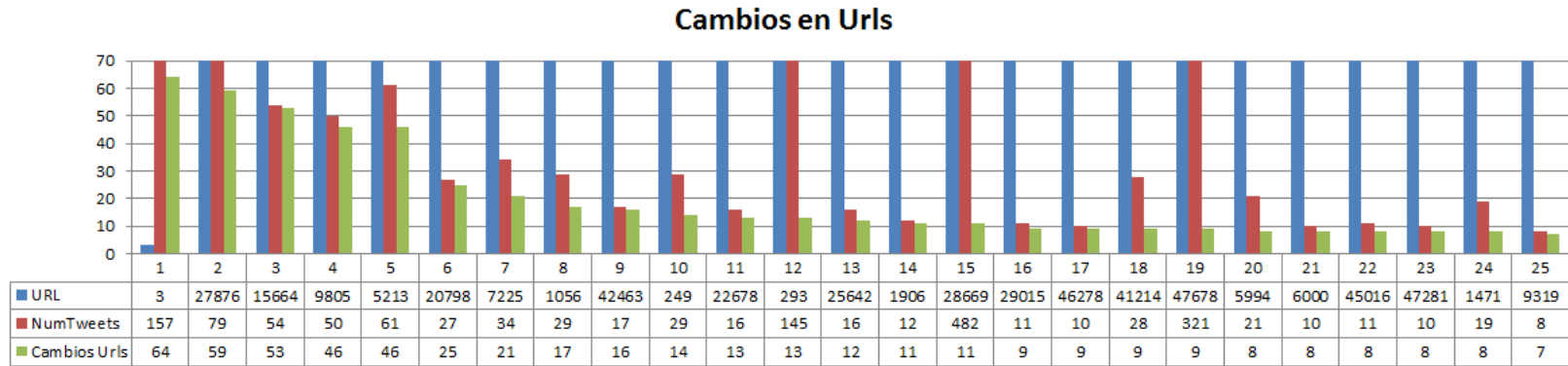


Figura 57: Urls con el mayor número de cambios en urls

En la figura se demuestra el resultado del menor número de cambios en urls que han compartido una misma url extendida.

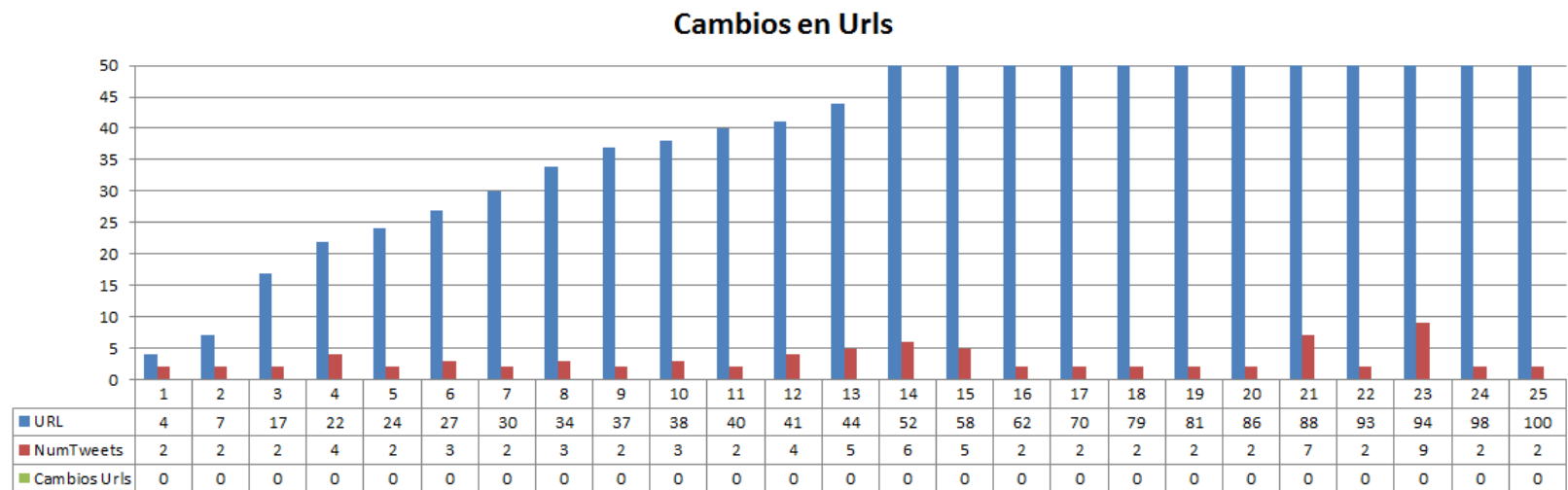


Figura 58: Urls con el menor número de cambios en urls

Los tweets que compartieron la URL 3 = [http://www.d.umn.edu/itss/training/online/webdesign/webdev\\_listserv.html#subscribe](http://www.d.umn.edu/itss/training/online/webdesign/webdev_listserv.html#subscribe), fueron los que mayor cambio tuvieron en las urls en donde existió 64 cambios de un total de 157 tweets, y la URL 4 = <http://lists.d.umn.edu/pipermail/webdev/2011q3/000598.html> que fue compartida en dos tweets no cambió las urls.

Se indica la información de la url con mayor y menor número de cambios en las urls.

Tabla 36: Url con mayor número de cambios en urls

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
3	157	4	21	64	25	28

Porcentaje del total de cambios del tweet con mayor cambio en urls.

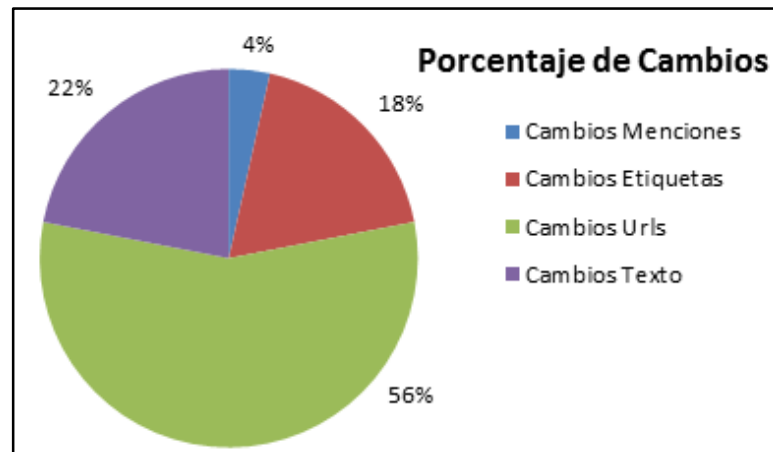


Figura 59: Porcentaje de la URL\_ID=3

Tabla 37: Url con menor número de cambios en urls

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
4	2	0	0	0	1	0

### 3.6. Urls con mayor y menor número de cambios en el texto

Se analiza el número de cambios que existe en el texto de los tweets que compartieron una misma url. Se muestra un resultado de 25 datos ordenados por “Cambios Texto” para determinar el mayor y menor número de cambios en esta métrica, como se lo indica en la siguiente tabla.

Tabla 38: Urls con mayor y menor número de cambios en el texto

Url con mayor número de cambios en Texto	URL	Cambios Texto	Url con menor número de cambios en Texto	URL	Cambios Texto
	2416	328		40	0
	28669	259		62	0
	28920	250		70	0
	36882	200		79	0
	31	183		86	0
	15053	162		114	0
	14156	155		142	0
	3365	146		158	0
	37205	139		170	0
	57304	131		174	0
	32853	130		182	0
	6743	126		183	0
	10183	122		184	0
	526	118		202	0
	30980	117		205	0
	41340	114		223	0
	33394	105		255	0
	44936	103		268	0
	13881	102		301	0
	2493	102		303	0
	40607	101		306	0
	47678	96		316	0
	17236	93		333	0
	166	92		340	0
5581	88	345	0		

En la figura se observa el resultado del mayor número de cambios en el texto que han compartido una misma url extendida.

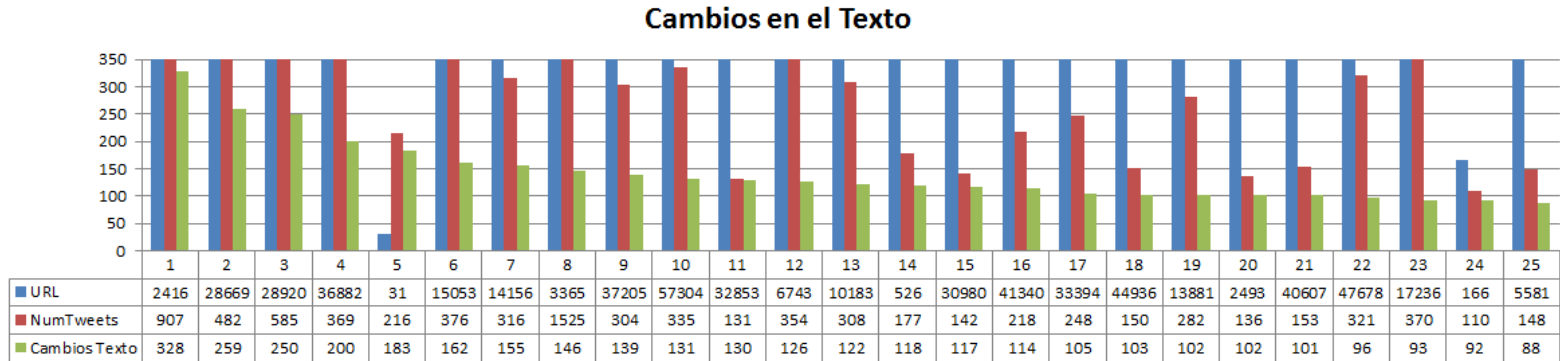


Figura 60: Urls con el mayor número de cambios en el texto

En la siguiente figura se indica el resultado del menor número de cambios en el texto que han compartido una misma url extendida.

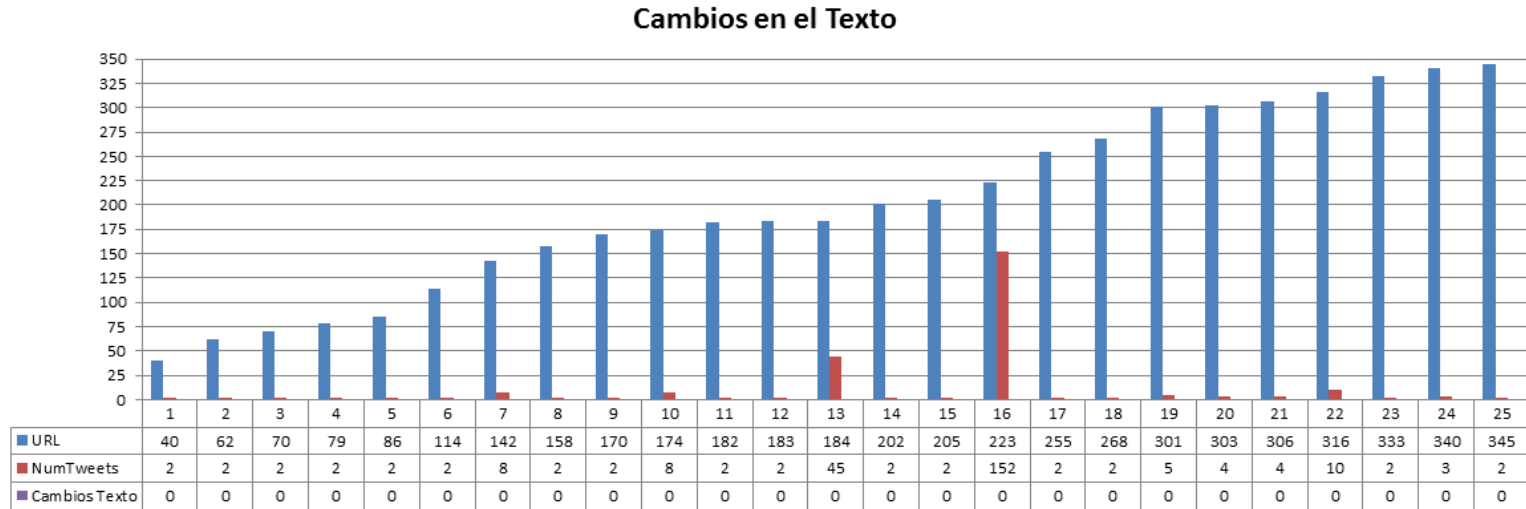


Figura 61: Urls con el menor número de cambios en el texto



Los tweets que compartieron la URL 2416= <http://www.petapixel.com/2011/11/29/cnn-lays-off-photojournalists-citing-the-accessibility-of-quality-cameras/>, fueron los que mayor cambio tuvieron con respecto al texto donde existió 328 cambios de un total de 907 tweets, y la URL 40= <http://owl.li/5yeZl> que fue compartida en dos tweets no cambió el texto.

A continuación en las siguientes tablas se indica la información de la url con mayor y menor número de cambios en el texto.

Tabla 39: Url con mayor número de cambios en el texto

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
2416	907	156	35	1	328	130

Porcentaje del total de cambios del tweet con mayor cambio en el texto.

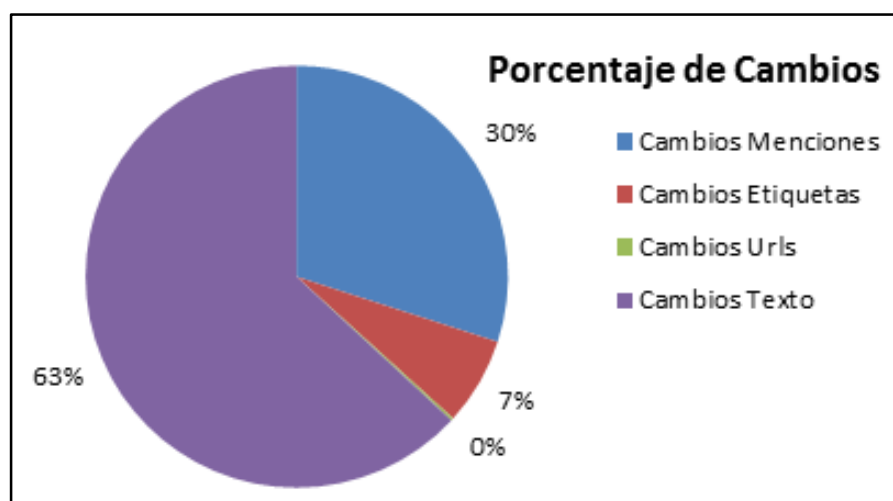


Figura 62: Porcentaje de cambios de la URL=2416

Tabla 40: Url con menor número de cambios en el texto

URL	NumTweets	Cambios Etiquetas	Cambios Menciones	Cambios Urls	Cambios Texto	Promedio Cambios
40	2	0	0	0	0	0

Los dos tweets que compartieron la URL=40 no presentaron ningún cambio en el texto, de igual manera las otras métricas de etiquetas, menciones, y urls de estos tweets no cambiaron, el usuario conservó el tweet en todos estos aspectos.

### 3.7. Urls con mayor y menor promedio en cambios

Finalizando con el estudio, se analiza el promedio de los cambios de las menciones, etiquetas, urls, y texto de los tweets que compartieron una misma url, se presenta un resultado con 25 datos que han sido ordenados por “Promedio Cambios” para determinar el mayor y menor número de cambios, como se lo indica en la siguiente tabla.

Tabla 41: Urls con mayor y menor número de promedio en cambios.

Url con mayor promedio en cambios	URL	Promedio Cambios	Url con menor promedio en cambios	URL	Promedio Cambios
	36882	140		40	0
	2416	130		62	0
	28669	117		70	0
	28920	115		79	0
	2582	114		86	0
	180	106		114	0
	276	101		142	0
	15053	76		158	0
	31	73		170	0
	14156	62		174	0
	10183	58		182	0
	13377	57		183	0
	1810	53		184	0
	6743	50		202	0
	57304	50		205	0
	3365	50		223	0
	44936	47		255	0
	41340	47		268	0
	526	47		301	0
	13881	46		303	0
	37205	44		306	0
	2493	42		316	0
	33394	42		333	0
	30980	42		340	0
40607	41	345	0		

Al aplicar el promedio de los cambios de una url, en la columna de menor promedio se tiene un resultado de 0 cambios lo que significa que los tweets que compartieron esas urls no cambiaron las etiquetas, menciones, urls y texto el usuario conservó el tweet en todos sus rasgos.

En la figura se muestra el resultado del mayor promedio en los cambios de los tweets que han compartido una misma url extendida.

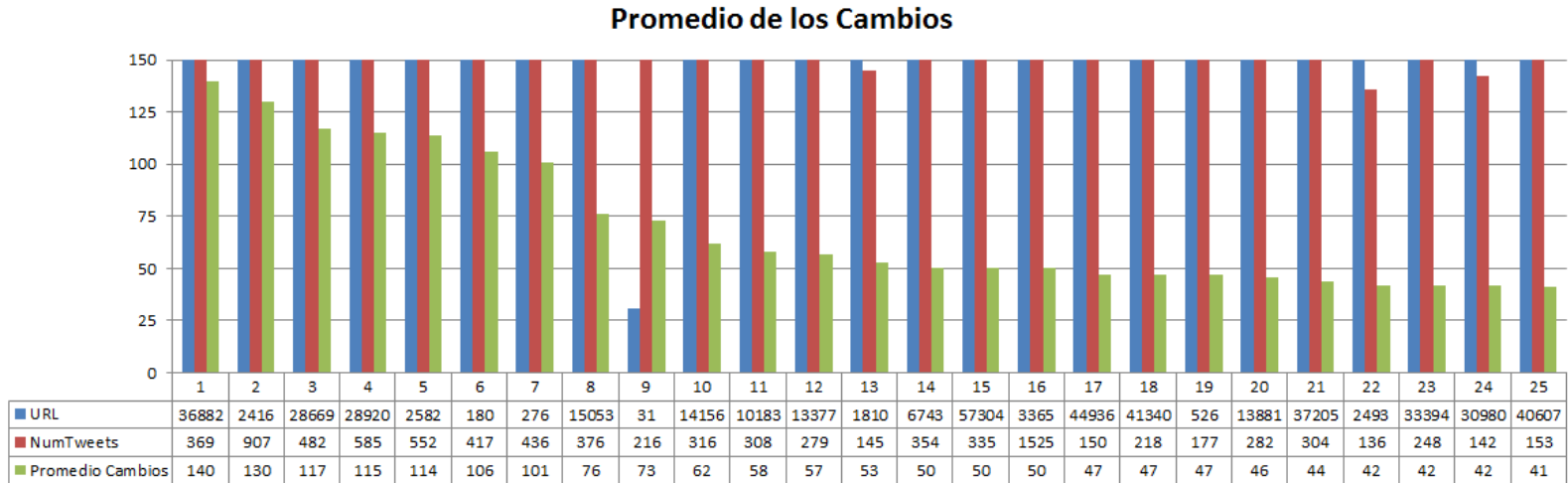


Figura 63: Urls con el mayor número de promedio de los cambios

En la figura se indica el resultado del menor promedio en los cambios de los tweets que compartieron una misma url extendida.

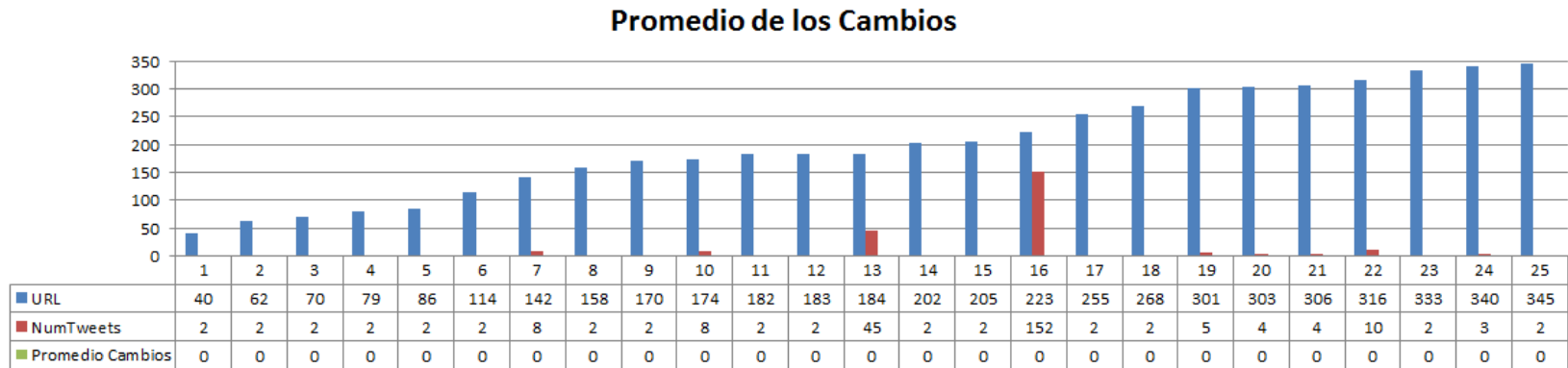


Figura 64: Urls con el menor número de promedio de los cambios

Los tweets que compartieron la URL 36882 = <http://www.ad-hoc-news.de/walmart-brings-affordability-accessibility-to-watkins--/en/corporate-news/23680899>, fueron los que mayor cambio tuvieron con respecto al promedio donde existió 140 cambios de un total de 369 tweets, y la URL 40= <http://owl.li/5yeZI> que fue compartida en dos tweets tiene el menor promedio.

A continuación se indica la información de la url con mayor y menor promedio de cambios.

Tabla 42: Url con mayor promedio en cambios

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
36882	369	360	0	1	200	140

Porcentaje del tweet con mayor promedio de cambios.

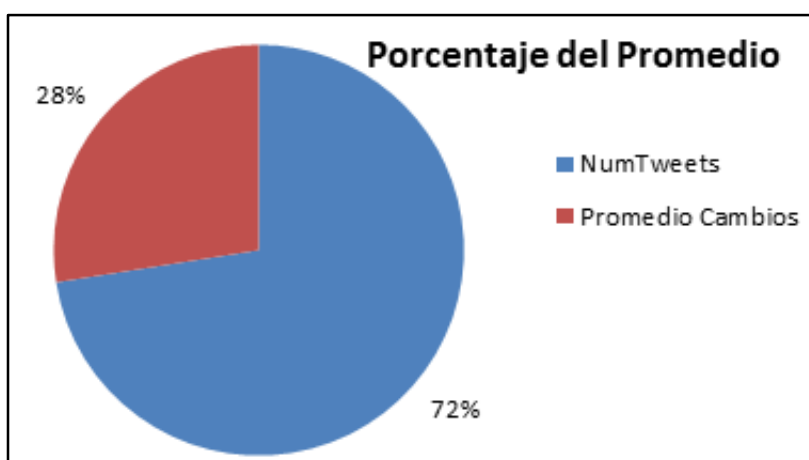


Figura 65: Promedio de cambios de la URL=36882

En la siguiente tabla consta la información de la url que tiene el menor promedio de los cambios en todas sus métricas.

Tabla 43: Url con menor promedio en cambios

URL	NumTweets	Cambios Menciones	Cambios Etiquetas	Cambios Urls	Cambios Texto	Promedio Cambios
40	2	0	0	0	0	0

Existe un total de 19534 urls que presentan al menos un cambio en cualquiera de sus métricas, mientras que las 2875 urls restantes que se compartieron en varios tweets no cambiaron en ninguna de sus métricas, lo que significa que los usuarios conservaron las menciones, etiquetas, urls y el texto igual que el tweet original.

Las 10 urls que mayor cambio tuvieron en el promedio de todas sus métricas se encuentran ordenadas por el atributo Promedio Cambios y son las siguientes:

Tabla 44: 10 urls con mayor cambio en el promedio de los tweets.

Promedio Cambios	URL_ID	URL
140	36882	<a href="http://www.homebuttonfix.com/2012/03/apple-home-button-fix.html">http://www.homebuttonfix.com/2012/03/apple-home-button-fix.html</a>
130	2416	<a href="http://www.petapixel.com/2011/11/29/cnn-lays-off-photojournalists-citing-the-accessibility-of-quality-cameras/">http://www.petapixel.com/2011/11/29/cnn-lays-off-photojournalists-citing-the-accessibility-of-quality-cameras/</a>
117	28669	<a href="http://www.mysqltalk.com/gaad.html">http://www.mysqltalk.com/gaad.html</a>
115	28920	<a href="http://www.mysqltalk.com/participate.html">http://www.mysqltalk.com/participate.html</a>
114	2582	<a href="http://paper.li/barrierfreemb/1321810432">http://paper.li/barrierfreemb/1321810432</a>
106	180	<a href="http://paper.li/umegaki/1313807124">http://paper.li/umegaki/1313807124</a>
101	276	<a href="http://paper.li/Hear_And_There/1306429577">http://paper.li/Hear_And_There/1306429577</a>
76	15053	<a href="https://www.facebook.com/globalaccessibilityawarenessday">https://www.facebook.com/globalaccessibilityawarenessday</a>
73	31	<a href="http://failsafe.fp.yahoo.com/404.html">http://failsafe.fp.yahoo.com/404.html</a>
62	14156	<a href="http://blogs.msdn.com/b/b8/archive/2012/02/14/enabling-accessibility.aspx">http://blogs.msdn.com/b/b8/archive/2012/02/14/enabling-accessibility.aspx</a>

Las 10 urls que menor cambio tuvieron en el promedio de todas sus métricas están ordenadas de acuerdo a Promedio Cambios y a la URL\_ID y son las siguientes:

Tabla 45: 10 urls con menor cambio en el promedio de los tweets.

URL_ID	URL
40	<a href="http://owl.li/5yeZl">http://owl.li/5yeZl</a>
62	<a href="http://www.youtube.com/watch?v=zUW8UhGX5AE">http://www.youtube.com/watch?v=zUW8UhGX5AE</a>
70	<a href="http://www.thaiwebaccessibility.com/apps/WebAccessibilitiy-consult.html">http://www.thaiwebaccessibility.com/apps/WebAccessibilitiy-consult.html</a>
79	<a href="http://www.mediaaccess.org.au/latest_news/australian-policy-and-legislation/government-health-information-made-more-accessible">http://www.mediaaccess.org.au/latest_news/australian-policy-and-legislation/government-health-information-made-more-accessible</a>
86	<a href="http://www.ozewai.org/users/spfeiffer">http://www.ozewai.org/users/spfeiffer</a>
114	<a href="http://www.amazon.com/Galvanized-Bleachers-Wheelchair-Accessibility-Footboards/dp/B0017D0UDQ?SubscriptionId=AKIAII366PKDKT2RZOPQ&amp;tag=fmtrn-20&amp;linkCode=sp1&amp;camp=2025&amp;creative=165953&amp;creativeASIN=B0017D0UDQ">http://www.amazon.com/Galvanized-Bleachers-Wheelchair-Accessibility-Footboards/dp/B0017D0UDQ?SubscriptionId=AKIAII366PKDKT2RZOPQ&amp;tag=fmtrn-20&amp;linkCode=sp1&amp;camp=2025&amp;creative=165953&amp;creativeASIN=B0017D0UDQ</a>
142	<a href="http://www.amazon.com/Galvanized-Bleachers-Wheelchair-Accessibility-Footboards/dp/B0017D28TA?SubscriptionId=AKIAJQ75T7T3QKXY4SNQ&amp;tag=neopt06-21&amp;linkCode=sp1&amp;camp=2025&amp;creative=165953&amp;creativeASIN=B0017D28TA&amp;">http://www.amazon.com/Galvanized-Bleachers-Wheelchair-Accessibility-Footboards/dp/B0017D28TA?SubscriptionId=AKIAJQ75T7T3QKXY4SNQ&amp;tag=neopt06-21&amp;linkCode=sp1&amp;camp=2025&amp;creative=165953&amp;creativeASIN=B0017D28TA&amp;</a>
158	<a href="http://kitchen.sirius-radio-reviews.com/microsoft-accessibility-instruction-sets-loose-the-effectiveness-of-the-pivot-table/?utm_source=twitterfeed&amp;utm_medium=twitter">http://kitchen.sirius-radio-reviews.com/microsoft-accessibility-instruction-sets-loose-the-effectiveness-of-the-pivot-table/?utm_source=twitterfeed&amp;utm_medium=twitter</a>
170	<a href="http://UXPrinciples.com">http://UXPrinciples.com</a>
174	<a href="http://howtohackthe.blogspot.com/2011/11/174-run-commands-for-windows-xp.html">http://howtohackthe.blogspot.com/2011/11/174-run-commands-for-windows-xp.html</a>

### 3.8. Promedio de los resultados en todas las métricas

En la siguiente tabla se promedia los 22409 resultados de cada métrica, se empieza con una muestra de 2500 datos y se va incrementando este valor hasta culminar con los mismos, como se puede observar la métrica que presenta el mayor número de cambios realizados es el texto, y la métrica que tiene el menor número de cambios son las urls.

Tabla 46: Promedio de los resultados de todas las métricas

Número de tweets	Promedios en Menciones	Promedios de Etiquetas	Promedio de Url	Promedios del Texto
2500	1,85314126	0,93917567	0,23569428	3,94037615
5000	1,41068214	0,78855771	0,19523905	3,45569114
7500	1,25950127	0,75290039	0,17362315	3,25923456
10000	1,12711271	0,70547055	0,15991599	3,05530553
12500	1,09544764	0,69757581	0,15297224	3,05696456
15000	1,07960531	0,68557904	0,13967598	3,02533502
17500	1,04468827	0,68861078	0,13726499	2,98697068
20000	1,02025101	0,6839342	0,13155658	2,93414671
22409	0,99669776	0,67044491	0,1258423	2,87424695

Como se lo puede observar del 100% el promedio de cambios de los 22409 tweets de acuerdo a cada métrica es el siguiente: Para el texto 62%, seguido de las menciones con un 21%, luego etiquetas con un 14%, y finalmente las urls que representan un 3%.

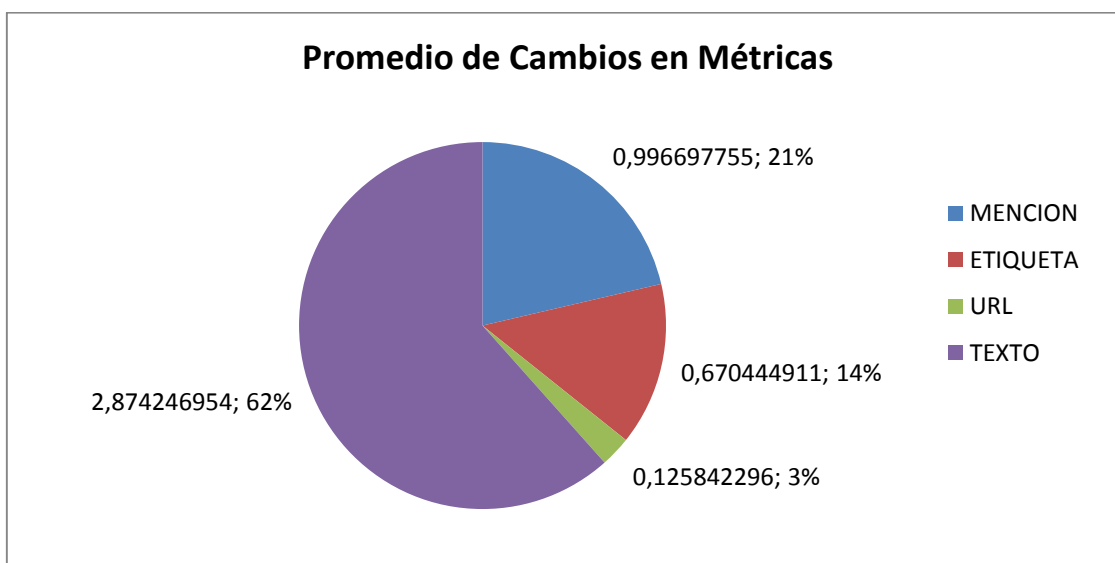


Figura 66: Porcentaje del promedio de cambios en todas las métricas

## CONCLUSIONES

Al haber culminado el proyecto, utilizando datos de twitter previamente ya recopilados, se aplicado un tratamiento de la información de los tweets para medir los cambios en las métricas planteadas y realizando un estudio real de los mismos se ha llegado a las siguientes conclusiones:

- Al analizar los tweets según las URLs que comparten, permite medir los cambios en las etiquetas, menciones, urls, y texto, para estudiar la difusión y la re-difusión de información en twitter.
- En el proceso de difusión de información, cuando los usuarios propagan un tweet lo que menos agregan al momento de retransmitirlo hacia sus seguidores son URLs, y lo que más añaden para enriquecer el tweet son etiquetas, texto y menciones.
- Los tweets que fueron más difundidos en la red son aquellos que contienen Urls, por tal razón se puede decir que los usuarios tienden a publicar enlaces que sirven como puerta de entrada a otras plataformas como: páginas web, blogs, fotografías o vídeos donde la información emitida en los 140 caracteres es desarrollada con mayor profundidad.
- Existe una relación entre los cambios de los tweets porque al observar los resultados se determina que si los usuarios cambian el texto también cambian las menciones, y al cambiar las menciones cambian las etiquetas, conservando en su mayoría las urls que comparten.
- Las urls que compartieron el mayor número de tweets y retweets se basan en facebook que es una red social y sobre el tema de la accesibilidad lo que significa que los usuarios le dieron mayor relevancia a estos tópicos al momento de twittear y retwittear tweets.

- Es conveniente aplicar métricas como: menciones, etiquetas, urls, y texto porque permiten analizar la forma de cómo se enriquecieron los tweets que compartieron una misma url y ver cuáles de estos han cambiado con mayor frecuencia en algunas de sus métricas antes mencionadas.
- Luego de haber analizado una cantidad de 22409 tweets, se observó que el mayor número de cambios realizados sobre un tweet es más notable con respecto al texto, seguida de las menciones, las etiquetas y en la que menor cambio se reflejó es en las urls, es decir, que a pesar de que las personas lleven a cabo modificaciones sobre el texto de un tweet tienden a conservar los rasgos clasificatorios más importantes como son las menciones, las etiquetas y las urls que están dentro de dicho tweet.



## RECOMENDACIONES

Una vez culminado el proyecto se presenta las siguientes recomendaciones:

Al trabajar con grandes volúmenes de datos es recomendable dividir el proceso en tareas más pequeñas que se ejecuten de manera independiente, con la finalidad de poder manejar toda la información y obtener los resultados esperados.

Para determinar un cambio real de los recursos se recomienda trabajar con las urls extendidas para recuperar los tweets que compartieron una misma url, debido que al seleccionar las url cortas hay un gran cambio en los resultados, el motivo es que existe url cortas diferentes que en su correspondiente url extendida se refieren a un mismo tema, esto se debe a que los usuarios utilizan diferentes acortadores de una url.

En la construcción del algoritmo, en la parte que se encarga de extraer datos se recomienda no implementar procesos complejos, así se logrará disminuir el tiempo de ejecución al momento de procesar grandes volúmenes de tweets.

Para realizar el procesamiento del texto se recomienda identificar los índices adecuados que permitan trabajar con la función `.indexOf()` de la clase `String` de Java, para extraer las menciones “@”, etiquetas “#”, urls “http://” y “https://”, con la finalidad de recuperar solamente el texto del tweet.

Analizar la información con la que se cuenta en la base de datos, para obtener datos precisos y poder realizar el tratamiento y manipulación de cada una de las métricas: menciones, etiquetas, urls, texto, evitando obtener datos innecesarios y sobre todo redundantes.

Utilizar medidas de dispersión para el análisis de datos porque las mismas permiten conocer si los valores en general están cerca o alejados de los valores centrales, muestran la variabilidad de una distribución de datos, obteniendo resultados más específicos y no generalizados.

## **TRABAJOS FUTUROS**

### **Tendencia de Tweets**

Se puede emplear el estudio de cambios en los tweets para determinar las tendencias que seguirán un conjunto específico de datos o usuarios, este trabajo emplea análisis cuantitativos para determinar los cambios llevados a cabo sobre un tweet; en el caso de que se quiera ampliar el campo de investigación de la interacción en redes sociales, se puede buscar los cambios realizados en una muestra de tweets, pero ya enmarcando el análisis sobre un campo en específico. Por ejemplo, se pueden llevar a cabo valoraciones y testeos de cambio en tweets que hablen sobre biología dentro de la UTPL, esto con el fin de establecer el auge y la influencia de un tema en específico dentro de una organización, empresa o comunidad virtual.

### **Recomendador de Recursos**

También mediante el análisis de cambios se pueden establecer las bases para diseñar, desarrollar e implementar un sistema recomendador o filtros de datos para búsqueda de información relevante dentro de una red social, enmarcándolo en un ámbito específico, en donde se pueda vincular términos como: etiquetas, menciones, y urls.

### **Análisis de Palabras**

En la parte del procesamiento del texto se propone normalizar el texto, trabajar el tema de las tildes, palabras que se escriben igual pero con diferente significado, analizar palabras en singular y plural, darle un tratamiento más avanzado enfocado en las palabras del tweet, con la finalidad de establecer y determinar coincidencias o relaciones de palabras.

### **Procesamiento del Lenguaje Natural**

Al tener ya un análisis cuantitativo de los cambios encontrados, se puede aplicar análisis de sentimientos o emociones de los usuarios identificando las opiniones que realizan respecto al nivel de aceptación positiva o negativa que toma respecto a un tema, en este caso tweets que comparten una misma url. Se puede explotar datos como: ubicación del usuario y el texto del tweet.

## BIBLIOGRAFÍA

- Rosado Hernández, M. (2 de Mayo de 2013). *Diseño de Programas. Pseudocódigo y Diagramas*. Obtenido de Ciclo Formativo de Grado Superior Administración de Sistemas Informáticos en Red: <http://informatica.iesvalledeljerteplasencia.es/wordpress/disenode-programas-pseudocodigo-y-diagramas/>
- Twitter. (2013). Obtenido de <http://www.twitter.com/>
- Twitter Developers. (13 de agosto de 2013). Obtenido de <https://dev.twitter.com/docs/platform-objects/tweets>
- Webyog. (2014). Obtenido de SQLyog MySQL GUI: <https://www.webyog.com/>
- Alcat, E. (2011). *¡Influye!: claves para dominar el arte de la persuasión*. Barcelona: Alienta. Obtenido de <http://www.iberlibro.com/Influye-Claves-dominar-arte-persuasi%C3%B3n-Enrique/8382294653/bd>
- Amblás, M., Bendito, E., Carmona, A., Encinas, A., Joan, J., & Mitjana, M. (2003). *Un procedimiento de difusión de la información*. Obtenido de <http://www.ma3.upc.edu/users/bencar/congresos/ivjmda-broad.pdf>
- Barcia Palacios, G. (Marzo de 2011). *¿Ruby on Rails como problema?* Obtenido de White Papers: <http://www.gerardobarcia.com/articulos/WhitepaperRoRcomoProblema.pdf>
- Carter, S., Tsagkias, M., & Weerkamp, W. (2011). *Twitter hashtags: Joint Translation and Clustering*. Obtenido de [http://www.websci11.org/fileadmin/websci/Posters/125\\_paper.pdf](http://www.websci11.org/fileadmin/websci/Posters/125_paper.pdf)
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). *Measuring User Influence in Twitter: The Million Follower Fallacy*. Obtenido de Max Planck Institute for Software Systems (MPI-SWS), Germany; Royal Veterinary College, University of London, United Kingdom; CS Dept., Federal University of Minas Gerais (UFMG), Brazil: <http://snap.stanford.edu/class/cs224w-readings/cha10influence.pdf>
- Chang, H.-C. (2010). *A New Perspective on Twitter Hashtag Use: Diffusion of Innovation Theory*. Obtenido de Department of Informatics, College of Computing and Information University at Albany, State University of New York: [http://mail.asis.org/asist2010/proceedings/proceedings/ASIST\\_AM10/submissions/295\\_Final\\_Submission.pdf](http://mail.asis.org/asist2010/proceedings/proceedings/ASIST_AM10/submissions/295_Final_Submission.pdf)
- Classora. (28 de Febrero de 2013). *Metadatos: definición, aplicaciones y estándares*. Obtenido de El Blog de Classora Líderes en volumen de datos y en tecnologías de

- enriquecimiento semántico: <http://blog.classora.com/2013/02/28/metadatos-definicion-aplicaciones-y-estandares/>
- Collado C, E. (2011). *Máster universitario de Sociedad de la Información y el Conocimiento*. Obtenido de <http://eduangi.com/wp-content/uploads/2011/07/El-microblogging-en-el-periodismo-esp%C3%B1ol.pdf>
- Congosto, M., Fernández, M., & Moro Egido, E. (2011). *Twitter y Política: Información, Opinión y ¿Predicción?* Obtenido de Cuadernos de Comunicación Evoca, 4 (Comunicación política 2.0), febrero, 10-15: <http://www.evocaimagen.com/cuadernos/cuadernos4.pdf>
- Dans, E. (17 de Agosto de 2011). *La generalización del uso de redes sociales*. Obtenido de El blog de Enrique Dans: <http://www.enriquedans.com/2011/08/la-generalizacion-del-uso-de-redes-sociales.html>
- De la Rosa Troyano, F. (Febrero de 2012). *Sistemas de Inteligencia Web: Análisis de Redes Sociales*. Obtenido de <http://www.lsi.us.es/~ffrosat/publicaciones/Tesis.2011.v75.vf.pdf>
- Domínguez, A. (5 de Diciembre de 2012). *Tutorial básico de Twitter*. Obtenido de Asociación Española para la Cultura, el Arte y la Educación (ASOCAE O.N.G.D.) - [www.asocae.org](http://www.asocae.org) - RNA: 592727 - CIF: G70195805: [http://www.cultureduca.com/blog/tutorial-basico-de-twitter/#lo\\_primer](http://www.cultureduca.com/blog/tutorial-basico-de-twitter/#lo_primer)
- Fabrega, J., & Paredes, P. (2013). *Social Contagion and Cascade Behaviors on Twitter*. Obtenido de ISSN 2078-2489: <http://www.mdpi.com/2078-2489/4/2/171>
- Gladwell, M. (4 de Octubre de 2010). *Small change: why the revolution not be tweeted*. Obtenido de The New Yorker : [http://www.newyorker.com/reporting/2010/10/04/101004fa\\_fact\\_gladwell?currentPage=all](http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell?currentPage=all)
- Gutiérrez-Rubí, A. (2014). *Twitter, mucho más que la CNN. La red social se ha convertido en el canal más libre, autónomo y dinámico para compartir información*. Obtenido de Artículos y reflexiones: <http://www.gutierrez-rubi.es/wp-content/uploads/2011/07/Twitter.pdf>
- Holmberg, K. (2012). *Online Information Ripples – A Conceptual Model for Analyzing Information Dissemination Patterns in Social media*. Obtenido de Department of Information Studies, Åbo Akademi University, Finland: [http://kimholmberg.fi/papers/holmberg\\_online\\_information\\_ripples\\_dgiconference2012.pdf](http://kimholmberg.fi/papers/holmberg_online_information_ripples_dgiconference2012.pdf)
- iLifebelt. (2013). *Estudio Redes Sociales Centroamérica 2013*. Obtenido de II Estudio de las Redes Sociales en América Central: <http://octavioislas.files.wordpress.com/2013/08/redes-sociales-centroamerica-2013.pdf>

- Iribarren, J. (17 de Julio de 2009). *Impact of Human Activity Patterns on the Dynamics of Information Diffusion*. Obtenido de IBM Corporation, ibm.com e-Relationship Marketing Europe, 28002 Madrid, Spain: <http://prl.aps.org/abstract/PRL/v103/i3/e038702>
- Jiménez, M. (2013). *Visualización de redes de colaboración a partir de datos abiertos en twitter aplicando técnica*. Obtenido de <http://dspace.utpl.edu.ec/bitstream/123456789/6595/1/Jimenez%20Mercy%20Elizabeth.pdf>
- Juárez, A. (2009). *Curso de Algoritmos*. Obtenido de Universidad Mariano Gálvez de Retalhuleu: <http://pakin.byethost22.com/wp-content/uploads/2009/10/lectura-2-tecnicas-de-representacion-algoritmica.pdf>
- Kawamoto, T. (25 de Septiembre de 2012). *A stochastic model of the tweet diffusion on the Twitter network*. Obtenido de Department of Physics, The University of Tokyo, Komaba, Meguro, Tokyo 153-8505, Japan: <http://arxiv.org/pdf/1209.5599.pdf>
- Kwak, H., Lee, C., Park, H., & Sue . (26-30 de abril de 2010). *What is Twitter, a Social Network or a News Media?* Obtenido de Department of Computer Science, KAIST - 335 Gwahangno, Yuseong-gu, Daejeon, Korea: <http://an.kaist.ac.kr/~haewoon/papers/2010-www-twitter.pdf>
- Lerman, K., & Ghosh, R. (2010). *Information Contagion: n Empirical Study of the*. Obtenido de Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media: <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1509/1839>
- López Abellán, M. (2012). *Twitter como instrumento de comunicación política en campaña: Elecciones Generales 2011*. Obtenido de Cuadernos de Gestión de Información. p.69-p.84. ISSN 2253-8429: <http://fcdmurcia.es/ojs/index.php?journal=cuadernos&page=article&op=download&path%5B%5D=107&path%5B%5D=30>.
- Lugo, M. D. (Enero de 2011). *Análisis de redes sociales en el mundo rural: guía inicial*. Obtenido de Revista de Estudios Sociales No. 38 rev.estud.soc. Pag.129-142: <http://res.uniandes.edu.co/view.php/681/view.php>
- Mejía, C. P. (Febrero de 2010). *Análisis de Redes Sociales a Gran Escala*. Obtenido de <http://www.cs.cinvestav.mx/TesisGraduados/2010/tesisCristianMejia.pdf>
- Meunier Rosas, F. (NOVIEMBRE de 2012). *El uso de la red social twitter como herramienta para la difusión de información pública*. Obtenido de Primera Revista Electrónica en América Latina Especializada en Comunicación: [http://www.razonypalabra.org.mx/N/N81/V81/27\\_Meunier\\_V81.pdf](http://www.razonypalabra.org.mx/N/N81/V81/27_Meunier_V81.pdf)

- Millán, A., Fermín, G., & Chacón, J. (13 de Mayo de 2008). *Proyecto de Carrera: Ingeniería Informática*. Obtenido de <http://www.monografias.com/trabajos59/diagrama-flujo/diagrama-flujo.shtml>
- Nikolov, S. (2012). *Information Diffusion at Twitter*. Obtenido de [http://blogs.ischool.berkeley.edu/i290-abdt-s12/files/2012/10/stan\\_diffusion\\_twitter.pdf](http://blogs.ischool.berkeley.edu/i290-abdt-s12/files/2012/10/stan_diffusion_twitter.pdf)
- O'Reilly, T., & Milstein, S. (2009). *The Twitter Book, Sebastopol: O'Reilly*. Obtenido de <http://www.cubanxgiants.com/berry/329/spring11/readings/week8/thetwitterbook.pdf>
- Orihuela, J. L. (2013). *Profesor en la Facultad de Comunicación de la Universidad de Navarra y autor del blog eCuaderno*. Obtenido de Una guía para comprender y dominar la plataforma que cambió la red. Alienta Editorial, Barcelona 2011.: <http://www.familyandmedia.eu/es/argumentos/facultades/238-guida-alla-comprensione-e-alluso-di-twitter.html>
- Orta, C., Pardo, F., & Salas, J. (2012). *Análisis de Redes de Influencia en Twitter*. Obtenido de [http://users.dsic.upv.es/grupos/nle/ceri/papers/ceri2012\\_guardian.pdf](http://users.dsic.upv.es/grupos/nle/ceri/papers/ceri2012_guardian.pdf)
- Ozdikis, O., Senkul, P., & Oguztuzun, H. ( 31 de August de 2012). *Semantic Expansion of Hashtags for Enhanced Event Detection in Twitter*. Obtenido de Middle East Technical University Ankara, Turkey: <http://www.cs.ubc.ca/~welu/woss2012/papers/1-ozdikis.pdf>
- Pérez Pérez, I., & Monsalve López, C. (2012). *Diseño de una propuesta de herramienta de diagramación para el desarrollo de algoritmos para computadora*. Obtenido de Universidad Autónoma del Estado de Hidalgo.: [http://www.uaeh.edu.mx/nuestro\\_alumnado/icbi/articulos/Diseno%20de%20una%20propuesta%20de%20herramienta%20de%20diagramacion.pdf](http://www.uaeh.edu.mx/nuestro_alumnado/icbi/articulos/Diseno%20de%20una%20propuesta%20de%20herramienta%20de%20diagramacion.pdf)
- Piscitelli, A. (2011). *Twitter, la revolución y los enfoques ni-ni*. En Orihuela, J.L. *Mundo Twitter* (p. 15-20). Barcelona: Alienta.
- Ponce, I. (17 de Abril de 2012). (17/4/2012): “ *Monográfico: Redes Sociales*”. *Observatorio Tecnológico NIPO:820-10-289-9*. Obtenido de <http://recursostic.educacion.es/observatorio/web/es/internet/web-20/1043-redes-sociales>
- Ramos Peinado , E. (20 de Diciembre de 2012). *Twitter*. Obtenido de Blog Historia de la Informática - Universidad Politécnica de Valencia: <http://histinf.blogs.upv.es/2012/12/20/twitter/>
- Rodríguez, R., & Ureña, D. (05 de 2012). *Ten reasons for Twitter's use like tool in political and electoral communication*. Obtenido de Universidad Pontificia de Salamanca.: [http://www.masconsulting.es/wp-content/uploads/2012/05/Art\\_ComunicacionPluralismo.pdf](http://www.masconsulting.es/wp-content/uploads/2012/05/Art_ComunicacionPluralismo.pdf)

- Romero, D., Meeder, B., & Kleinberg, J. (Junio de 2011). *"Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter"*, (ed.) *International World Wide Web Conference Committee (IW3C2)*. Obtenido de <http://www.cs.cornell.edu/home/kleinber/www11-hashtags.pdf>
- Sanz, L. (Junio de 2003). *Análisis de Redes Sociales: o como representar las estructuras sociales subyacentes*. Obtenido de <http://digital.csic.es/bitstream/10261/1569/1/dt-0307.pdf>
- Suh, B., Hong, L., Pirolli, P., & Ed, H. (25 de 06 de 2010). *Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network*. Obtenido de IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust: <http://www-users.cs.umn.edu/~echi/papers/2010-socialcom/2010-06-25-retweetability-cameraready-v3.pdf>
- T2O. (10 de Enero de 2010). *AdMedia Services S.L.* Obtenido de Origen de Twitter: <http://www.t2o.es/blog/social-media/origen-de-twitter/Origen de Twitter>
- Thomases, H. (2010). *Twitter Marketing*. Indianapolis: Wiley.
- Twitter. (2013). *El glosario de Twitter*. Obtenido de <https://support.twitter.com/articles/352810-el-glosario-de-twitter#>
- Vazquez Gomez, J. B. (2012). *Análisis y Diseño de Algoritmos*. Obtenido de Red tercer milenio: [http://www.aliatuniversidades.com.mx/bibliotecasdigitales/pdf/sistemas/Analisis\\_y\\_diseno\\_de\\_algoritmos.pdf](http://www.aliatuniversidades.com.mx/bibliotecasdigitales/pdf/sistemas/Analisis_y_diseno_de_algoritmos.pdf)
- Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., y otros. (20 de Junio de 2013). *The Role of Information Diffusion*. Obtenido de <http://arxiv.org/pdf/1302.6276v2.pdf>
- Weng, Lilian; Menczer, Filippo; Ahn, Yong Yeol. (28 de August de 2013). *Virality Prediction and Community Structure in Social Networks*. Obtenido de Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA.: <http://www.nature.com/srep/2013/130828/srep02522/pdf/srep02522.pdf>
- Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., y otros. (October de 2010). *Understanding Retweeting Behaviors in Social Networks*. Obtenido de Tsinghua National Laboratory for Information Science and Technology: <http://people.cs.umass.edu/~jingyi/CIKM10-Yang-et-al-Understanding-Retweeting.pdf>
- Zhao, D., & Rosson, M. (2009). *How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication at Work*. Obtenido de Pennsylvania State University:

[http://www.pensivepuffin.com/dwmcphd/syllabi/info447\\_au10/readings/zhao.rosson.Twitter.GROUP09.pdf](http://www.pensivepuffin.com/dwmcphd/syllabi/info447_au10/readings/zhao.rosson.Twitter.GROUP09.pdf)



# ANEXOS

## 1. APLICACIÓN

### 1.1. DIAGRAMA DE CLASES

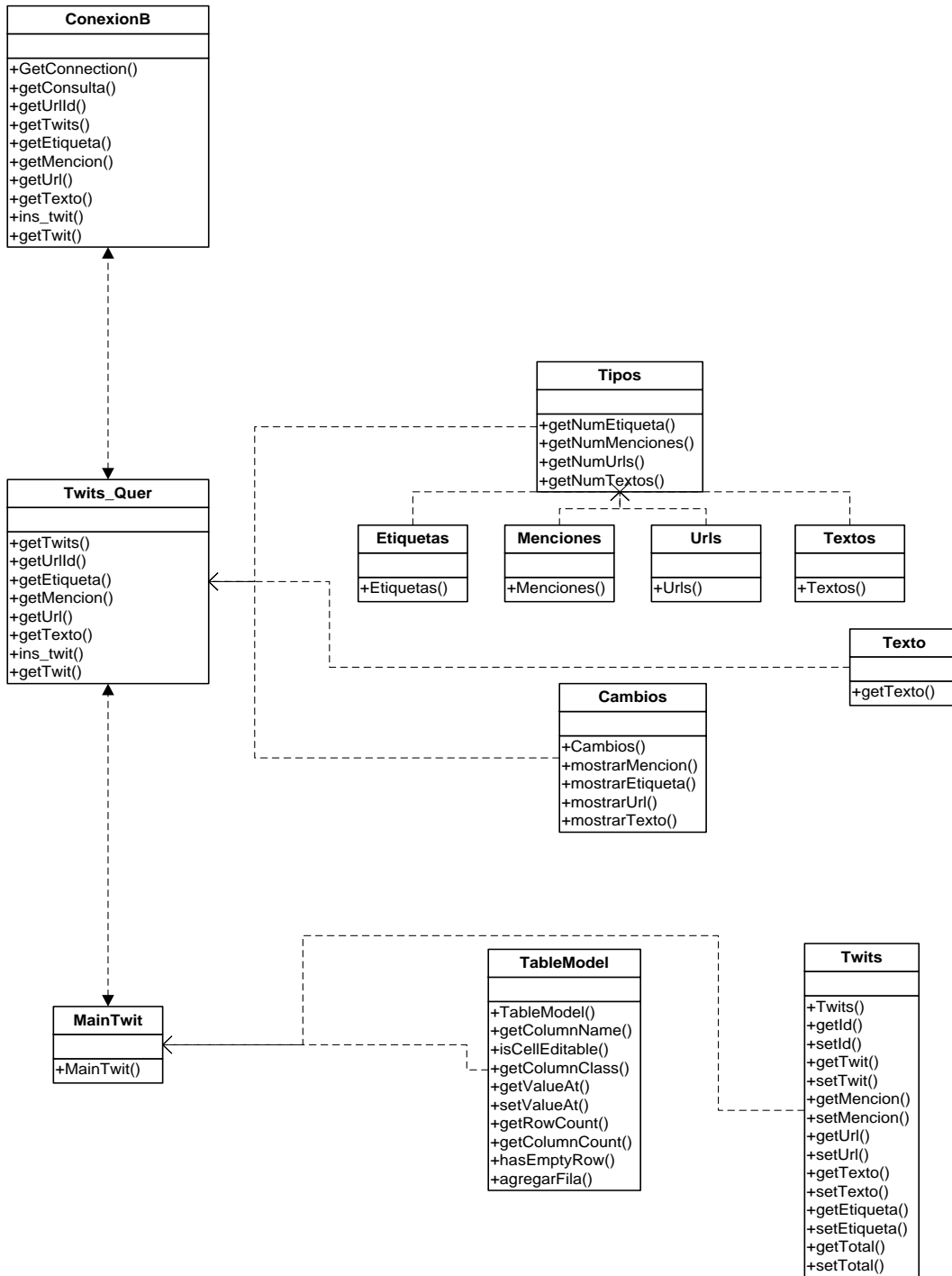


Figura 67: Diagrama de clases

## 1.2. MODELO RELACIONAL DE LA BASE DE DATOS SNA\_TWITTER.

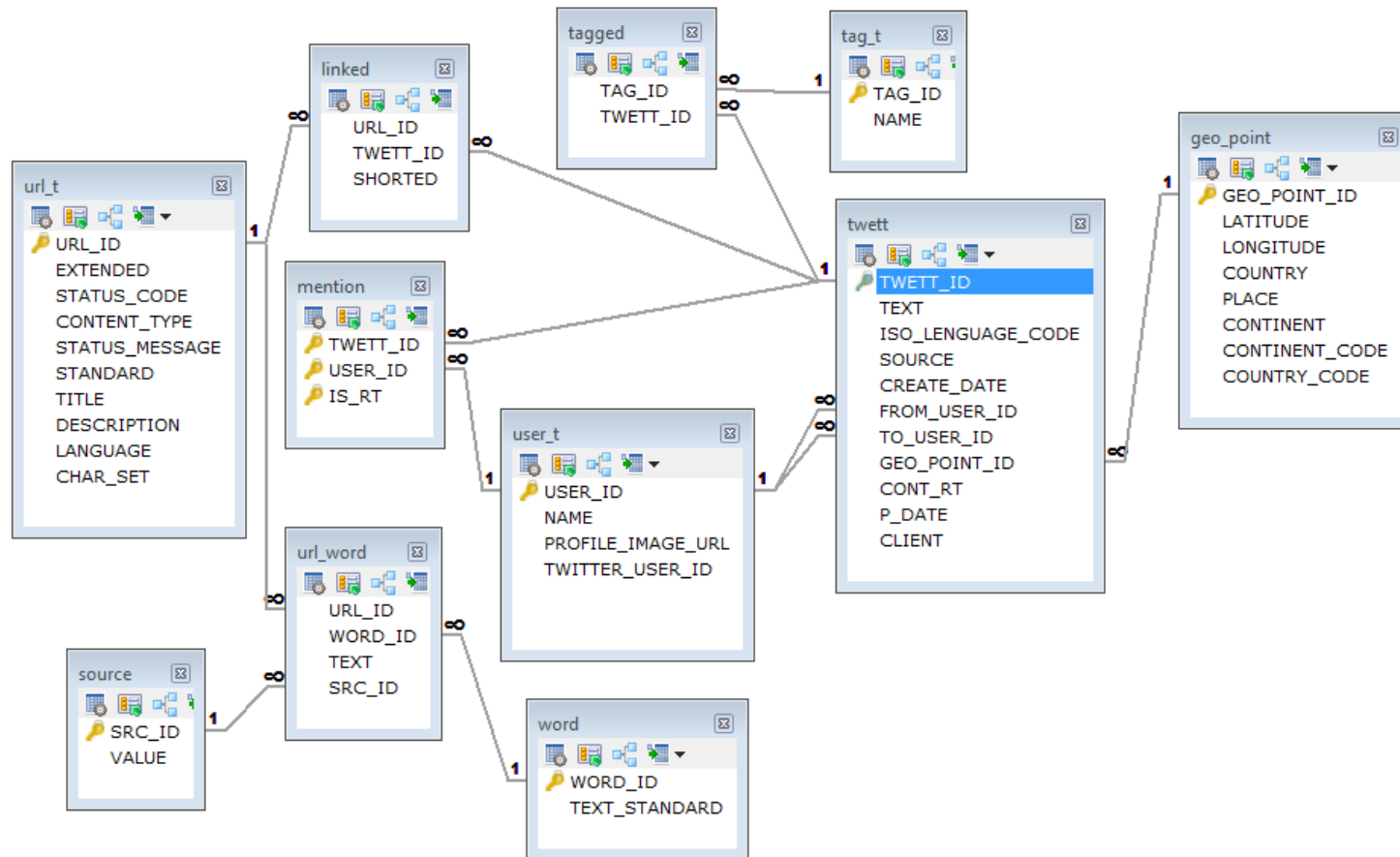


Figura 68: Modelo relacional de la base de datos

## 2. Stored Procedure

### 2.1. Concatenar menciones

```
DELIMITER $$

DROP PROCEDURE IF EXISTS `gp`.`prc_obt_mencion`$$

CREATE DEFINER=`root`@`localhost` PROCEDURE `prc_obt_mencion`(IN id
INT , OUT mencion VARCHAR(1024))
BEGIN
    DECLARE _EXTENDED VARCHAR(5120);
    DECLARE _TEXT VARCHAR(512);
    DECLARE _NAME VARCHAR(256);
    DECLARE salir BOOLEAN;
    DECLARE c_mencion CURSOR FOR SELECT u.EXTENDED, t.TEXT, s.NAME
        FROM url_t u, linked l, twett t, mention m, user_t s
        WHERE u.URL_ID = l.URL_ID
        AND l.TWETT_ID = t.TWETT_ID
        AND t.TWETT_ID = id
        AND t.TWETT_ID = m.TWETT_ID
        AND m.USER_ID = s.USER_ID;

    DECLARE CONTINUE HANDLER FOR NOT FOUND SET salir = TRUE;
    SET mencion= CONCAT('','');
    OPEN c_mencion;

    mencion_loop: LOOP
        FETCH c_mencion INTO _EXTENDED, _TEXT, _NAME;

        IF salir THEN
            CLOSE c_mencion;
            LEAVE mencion_loop;
        END IF;
        IF mencion='' THEN
            SET mencion= CONCAT(mencion, _NAME);
        ELSE
            SET mencion= CONCAT(mencion, ', ', _NAME);
        END IF;
    END LOOP mencion_loop;
END$$

DELIMITER ;
```

## 2.2. Concatenar etiquetas

```
DELIMITER $$

DROP PROCEDURE IF EXISTS `gp`.`prc_obt_etiqueta`$$

CREATE DEFINER=`root`@`localhost` PROCEDURE `prc_obt_etiqueta`(IN id
INT , OUT etiquetas VARCHAR(1024))
BEGIN
    DECLARE _EXTENDED VARCHAR(5120);
    DECLARE _TEXT VARCHAR(512);
    DECLARE _NAME VARCHAR(256);
    DECLARE salir BOOLEAN;
    DECLARE c_etiqueta CURSOR FOR SELECT u.EXTENDED, t.TEXT, n.NAME
        FROM url_t u, linked l, twett t, tagged a, tag_t n
        WHERE u.URL_ID = l.URL_ID
        AND l.TWETT_ID = t.TWETT_ID
        AND t.TWETT_ID = a.TWETT_ID
        AND a.TAG_ID = n.TAG_ID
        AND t.TWETT_ID = id;

    DECLARE CONTINUE HANDLER FOR NOT FOUND SET salir = TRUE;
    SET etiquetas= CONCAT('','');
    OPEN c_etiqueta;

    etiqueta_loop: LOOP
    FETCH c_etiqueta INTO _EXTENDED, _TEXT, _NAME;

    IF salir THEN
        CLOSE c_etiqueta;
        LEAVE etiqueta_loop;
    END IF;
    IF etiquetas='' THEN
        SET etiquetas= CONCAT(etiquetas,_NAME);
    ELSE
        SET etiquetas= CONCAT(etiquetas,',',_NAME);
    END IF;
    END LOOP etiqueta_loop;

END$$

DELIMITER ;
```

### 2.3. Concatenar urls

```
DELIMITER $$

DROP PROCEDURE IF EXISTS `gp`.`prc_obt_url`$$

CREATE DEFINER=`root`@`localhost` PROCEDURE `prc_obt_url`(IN id INT ,
OUT urls VARCHAR(1024))
BEGIN
    DECLARE _EXTENDED VARCHAR(5120);
    DECLARE _TEXT VARCHAR(512);
    DECLARE _NAME VARCHAR(256);
    DECLARE salir BOOLEAN;
    DECLARE c_url CURSOR FOR SELECT u.EXTENDED, t.TEXT, L.SHORTED
        FROM linked l, twett t, url_t u
        WHERE T.TWETT_ID = L.TWETT_ID
        AND u.URL_ID = l.URL_ID
        AND t.TWETT_ID = id;

    DECLARE CONTINUE HANDLER FOR NOT FOUND SET salir = TRUE;
    SET urls= CONCAT('','');
    OPEN c_url;

    url_loop: LOOP
    FETCH c_url INTO _EXTENDED, _TEXT, _NAME;

    IF salir THEN
        CLOSE c_url;
        LEAVE url_loop;
    END IF;
    IF urls='' THEN
        SET urls= CONCAT(urls,_EXTENDED);
    ELSE
        SET urls= CONCAT(urls,',',_EXTENDED);
    END IF;
    END LOOP url_loop;

END$$

DELIMITER ;
```

## 2.4. Concatenar tweets

```
DELIMITER $$

DROP PROCEDURE IF EXISTS `gp`.`prc_obt_texto`$$

CREATE DEFINER=`root`@`localhost` PROCEDURE `prc_obt_texto`(IN id INT
, OUT textos VARCHAR(1540))
BEGIN
    DECLARE _TEXT VARCHAR(512);
    DECLARE salir BOOLEAN;
    DECLARE c_textos CURSOR FOR SELECT DISTINCT t.TEXT
        FROM twett t, linked l, url_t u
        WHERE t.TWETT_ID = l.TWETT_ID
            AND l.URL_ID = u.URL_ID
            AND t.TWETT_ID = id;

    DECLARE CONTINUE HANDLER FOR NOT FOUND SET salir = TRUE;
    SET textos= CONCAT('', '');
    OPEN c_textos;

    texto_loop: LOOP
        FETCH c_textos INTO _TEXT;

        IF salir THEN
            CLOSE c_textos;
            LEAVE texto_loop;
        END IF;
        IF textos='' THEN
            SET textos= CONCAT(textos,_TEXT);
        ELSE
            SET textos= CONCAT(textos,',',_TEXT);
        END IF;
    END LOOP texto_loop;

END$$

DELIMITER ;
```

## 2.5. Insertar resultados

```
DELIMITER $$

DROP PROCEDURE IF EXISTS `gp`.`prc_ins_tweet`$$

CREATE DEFINER=`root`@`localhost` PROCEDURE `prc_ins_tweet`(IN url_id
INT , IN num_tweet INT, IN mencion INT, IN etiqueta INT, IN url INT ,
IN texto INT, IN total INT)
BEGIN

    INSERT INTO `gp_resultados` (`gp_url_id`, `gp_num_tweet`,
`gp_mencion`, `gp_etiqueta`, `gp_url`, `gp_texto`, `gp_total`)
    VALUES(url_id, num_tweet, mencion,etiqueta,url,texto, total);
COMMIT;
END$$

DELIMITER ;
```

## 3. Clase Base

```
import java.sql.*;
```

### 3.1. Conexión a la base de datos

```
public class ConexionBD {
    private static Connection conexion = null;

    public static Connection GetConnection() {
        try {
            if (conexion == null) {
                Class.forName("com.mysql.jdbc.Driver");
                String url = "jdbc:mysql://localhost:3306/gp";
                conexion = DriverManager.getConnection(url, "root", "");
            }
            return conexion;
        } catch (Exception ex) {
            System.out.println("ERROR " + ex);
        }
        finally
        {
            return conexion;
        }
    }
}
```

### 3.2. Obtener urls

```
public static ResultSet getUrlId() throws SQLException {
    String consulta = " SELECT u.URL_ID, count(*) num "
```

```

+ "FROM twett t, linked l, url_t u "
+ "WHERE t.TWETT_ID = l.TWETT_ID "
+ "AND l.URL_ID = u.URL_ID "
+ "GROUP BY u.EXTENDED "
+ "HAVING COUNT(*) > 1 "
+ "ORDER BY u.URL_ID "
+ "LIMIT 0, 22409";

```

```

ResultSet rs = null;
Connection conexion = ConexionBD.GetConnection();
if (conexion != null) {
    Statement s = conexion.createStatement();
    rs = s.executeQuery(consulta);
}
return rs;
}

```

### 3.3. Obtener tweet

```

public static ResultSet getTwits(int urlId) throws SQLException {
    String consulta = "SELECT DISTINCT t.TWETT_ID "
        + "FROM twett t, linked l "
        + " WHERE t.TWETT_ID = l.TWETT_ID "
        + " AND l.URL_ID =" + urlId;
    ResultSet rs = null;
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        Statement s = conexion.createStatement();
        rs = s.executeQuery(consulta);
    }
    return rs;
}

```

### 3.4. Obtener Etiqueta

```

public static String getEtiqueta(int twit) throws SQLException {

    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        CallableStatement prcEtiqueta = conexion.prepareCall("{call gp.prc_obt_etiqueta(?,?)}");
        //Parametro de entrada del procedimiento
        prcEtiqueta.setInt("id", twit);
        //Parámetro de salida del procedimiento
        prcEtiqueta.registerOutParameter("etiquetas", java.sql.Types.VARCHAR);
        //Ejecución del procedimiento
        prcEtiqueta.execute();
        //Obtención del parámetro de salida procedimiento prc_obt_etiqueta - etiquetas
        return prcEtiqueta.getString(2);
    }
    return "";
}

```



### 3.5. Obtener Mención

```
public static String getMencion(int twit) throws SQLException {
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        //Crear invocación del procedimiento prc_obt_mencion
        CallableStatement prcMencion=conexion.prepareCall("{call gp.prc_obt_mencion(?,?)}");
        //Parámetro de entrada del procedimiento
        prcMencion.setInt("id", twit);
        //Parámetro de salida del procedimiento
        prcMencion.registerOutParameter("mencion", java.sql.Types.VARCHAR);
        //Ejecución del procedimiento
        prcMencion.execute();
        //Obtención del parámetro de salida menciones
        return prcMencion.getString(2);
    }
    return "";
}
```

### 3.6. Obtener Url

```
public static String getUrl(int twit) throws SQLException {
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        //Crear invocación del procedimiento prc_obt_url
        CallableStatement prcUrl = conexion.prepareCall("{call gp.prc_obt_url(?,?)}");
        //Parámetro de entrada del procedimiento
        prcUrl.setInt("id", twit);
        //Parámetro de salida del procedimiento
        prcUrl.registerOutParameter("urls", java.sql.Types.VARCHAR);
        //Ejecución del procedimiento
        prcUrl.execute();
        //Obtención del parámetro de salida procedimiento prc_obt_url - url
        return prcUrl.getString(2);
    }
    return "";
}
```

### 3.7. Obtener Texto

```
public static String getTexto(int twit) throws SQLException {
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        //Crear invocación del procedimiento prc_obt_texto
        CallableStatement prcTexto = conexion.prepareCall("{call gp.prc_obt_texto(?,?)}");
        //Parámetro de entrada del procedimiento
        prcTexto.setInt("id", twit);
        //Parámetro de salida del procedimiento
        prcTexto.registerOutParameter("textos", java.sql.Types.VARCHAR);
        //Ejecución del procedimiento
```

```

        prcTexto.execute();
        //Obtención del parámetro de salida procedimiento prc_obt_texto - texto
        return prcTexto.getString(2);
    }
    return "";
}

```

### 3.8. Guardar resultados

```

public static void ins_tweet(int url_id, int num_tweet, int mencion, int etiqueta, int url, int texto, int
total) throws SQLException {
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        CallableStatement prcTweet = conexion.prepareCall("{call gp.prc_ins_tweet(?,?,?,?,?,?)}");
        prcTweet.setInt("url_id", url_id);
        prcTweet.setInt("num_tweet", num_tweet);
        prcTweet.setInt("mencion", mencion);
        prcTweet.setInt("etiqueta", etiqueta);
        prcTweet.setInt("url", url);
        prcTweet.setInt("texto", texto);
        prcTweet.setInt("total", total);
        prcTweet.execute();
    }
}

```

### 3.9. Almacenar los resultados

```

public static void ins_tweet(int url_id, int num_tweet, int mencion, int etiqueta, int url, int texto, int
total) throws SQLException {

    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        CallableStatement prcTweet = conexion.prepareCall("{call gp.prc_ins_tweet(?,?,?,?,?,?)}");
        prcTweet.setInt("url_id", url_id);
        prcTweet.setInt("num_tweet", num_tweet);
        prcTweet.setInt("mencion", mencion);
        prcTweet.setInt("etiqueta", etiqueta);
        prcTweet.setInt("url", url);
        prcTweet.setInt("texto", texto);
        prcTweet.setInt("total", total);
        prcTweet.execute();
    }
}

```

### 3.10. Obtener Resultados

```

public static ResultSet getTweet() throws SQLException {
    String consulta = " SELECT gp_url_id, "
        + "gp_num_tweet, "
        + "gp_mencion, "

```

```

        + "gp_etiqueta, "
        + "gp_url, "
        + "gp_texto "
        + "FROM gp.gp_resultados ";
    ResultSet rs = null;
    Connection conexion = ConexionBD.GetConnection();
    if (conexion != null) {
        Statement s = conexion.createStatement();
        rs = s.executeQuery(consulta);
    }
    return rs;
}
}

```

#### 4. CLASE Twits\_Query

```

import Base.ConexionBD;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.util.logging.Level;
import java.util.logging.Logger;

public class Twits_Query {
    public ResultSet getTwits(int urlId)
    {
        try {
            return ConexionBD.getTwits(urlId);
        } catch (SQLException ex) {
            Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
        }
        return null;
    }

    public ResultSet getUrlId()
    {
        try {
            return ConexionBD.getUrlId();
        } catch (SQLException ex) {
            Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
        }
        return null;
    }

    public String getEtiqueta(int twit ){
        try {
            return ConexionBD.getEtiqueta(twit);
        } catch (SQLException ex) {
            Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
        }
        return null;
    }
}

```

```

}

public String getMencion(int twit)
{
    try {
        return ConexionBD.getMencion(twit);
    } catch (SQLException ex) {
        Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
    }
    return null;
}

public String getUrl(int twit)
{
    try {
        return ConexionBD.getUrl(twit);
    } catch (SQLException ex) {
        Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
    }
    return null;
}

public String getTexto(int twit)
{
    try {
        return ConexionBD.getTexto(twit);
    } catch (SQLException ex) {
        Logger.getLogger(Twits_Query.class.getName()).log(Level.SEVERE, null, ex);
    }
    return null;
}

public void ins_twit(int url_id, int num_tweet, int mencion, int etiqueta, int url, int texto, int
total ) throws SQLException
{
    ConexionBD.ins_twit(url_id, num_tweet, mencion, etiqueta, url, texto, total);
}

public static ResultSet getTwit() throws SQLException
{
    return ConexionBD.getTwit();
}
}

```

## 5. Clase Tipos

```

import Logica.Etiquetas;
import java.util.ArrayList;

public class Tipos

```

```
{
```

### 5.1. Número de repeticiones de las etiquetas de los tweets.

```
public ArrayList<Cambios> getNumEtiqueta(ArrayList etiqueta, ArrayList etiqueta_aux)
{
    int _nVeces=0;
    //Array del objeto Cambios
    ArrayList<Cambios> rep=new ArrayList<Cambios>();

    int _tamanio=etiqueta_aux.size();
    //Para realizar comparaciones del objeto etiqueta y etiqueta_aux
    for (int i = 0; i < etiqueta.size(); i++) {
        _tamanio=etiqueta_aux.size();
        //Obtiene del arraylist y realiza casting al objeto Etiquetas
        Etiquetas str=(Etiquetas) etiqueta.get(i);
        for (int j = 0; j < _tamanio; j++) {
            Etiquetas str_=(Etiquetas) etiqueta.get(j);
            if( str.etiqueta.compareTo(str_.etiqueta)==0)
                {
                    _nVeces++;
                    etiqueta_aux.remove(j);
                    _tamanio=etiqueta_aux.size();
                    j=j-1;
                }
        }
        //Agrega ítem al objeto rep y crea una instancia de la clase Cambios
        rep.add(new Cambios(_nVeces,str.etiqueta));
        _nVeces=0;
        i=i-1;
    }
    return rep;
}
```

### 5.2. Número de repeticiones de las menciones de los tweets.

```
public ArrayList<Cambios> getNumMenciones(ArrayList mencion, ArrayList mencion_aux)
{
    int _nVeces=0;
    //Array del objeto Cambios
    ArrayList<Cambios> rep=new ArrayList<Cambios>();

    int _tamanio=mencion_aux.size();
    //Para realizar comparaciones del objeto mencion y mencion_aux
    for (int i = 0; i < mencion.size(); i++) {
        _tamanio=mencion_aux.size();
        //Obtiene del arraylist y realiza casting al objeto Menciones
        Menciones str=(Menciones) mencion.get(i);
        for (int j = 0; j < _tamanio; j++) {
            Menciones str_=(Menciones) mencion.get(j);
```

```

        if( str.mencion.compareTo(str_.mencion)==0)
            {
            _nVeces++;
            mencion_aux.remove(j);
            _tamanio=mencion_aux.size();
            j=j-1;
            }
        }
//Agrega ítem al objeto rep y crea una instancia de la clase Cambios
rep.add(new Cambios(_nVeces,str.mencion));
_nVeces=0;
i=i-1;
}
return rep;
}

```

### 5.3. Número de repeticiones de las urls de los tweets.

```

public ArrayList<Cambios> getNumUrls(ArrayList url, ArrayList url_aux)
{
int _nVeces=0;
//Array del objeto Cambios
ArrayList<Cambios> rep=new ArrayList<Cambios>();

int _tamanio=url_aux.size();
//Para realizar comparaciones del objeto url y url_aux
for (int i = 0; i < url.size(); i++) {
    _tamanio=url_aux.size();
    //Obtiene del arraylist y realiza casting al objeto Urls
    Urls str=(Urls) url.get(i);
    for (int j = 0; j < _tamanio; j++) {
        Urls str_=(Urls) url.get(j);
        if( str.url.compareTo(str_.url)==0)
            {
            _nVeces++;
            url_aux.remove(j);
            _tamanio=url_aux.size();
            j=j-1;
            }
        }
//Agrega ítem al objeto rep y crea una instancia de la clase Cambios
rep.add(new Cambios(_nVeces,str.url));
_nVeces=0;
i=i-1;
}
return rep;
}

```

### 5.4. Número de repeticiones del texto de los tweets.

```

public ArrayList<Cambios> getNumTextos(ArrayList texto, ArrayList texto_aux)
{
    int _nVeces=0;
    //Array del objeto Cambios
    ArrayList<Cambios> rep=new ArrayList<Cambios>();

        int _tamanio=texto_aux.size();
        //Para realizar comparaciones del objeto texto y texto_aux
        for (int i = 0; i < texto.size(); i++) {
            _tamanio=texto_aux.size();
            //Obtiene del arraylist y realiza casting al objeto Textos
            Textos str=(Textos) texto.get(i);
            for (int j = 0; j < _tamanio; j++) {
                Textos str_=(Textos) texto.get(j);
                if( str.texto.compareTo(str_.texto)==0)
                    {
                        _nVeces++;
                        texto_aux.remove(j);
                        _tamanio=texto_aux.size();
                        j=j-1;
                    }
            }
            //Agrega ítem al objeto rep y crea una instancia de la clase Cambios
            rep.add(new Cambios(_nVeces,str.texto));
            _nVeces=0;
            i=i-1;
        }
        return rep;
    }
}

```

## 6. Clase Etiquetas

```

public class Etiquetas {

    public int etiquetald;
    public String etiqueta;

    public Etiquetas(int id, String data)
    {
        etiquetald=id;
        etiqueta=data;
    }
}

```

## 7. Clase Menciones

```
public class Menciones {  
    public int mencionId;  
    public String mencion;  
  
    public Menciones(int id, String data)  
    {  
        mencionId=id;  
        mencion=data;  
    }  
}
```

## 8. Clase Textos

```
public class Textos {  
    public int textold;  
    public String texto;  
  
    public Textos(int id, String data)  
    {  
        textold=id;  
        texto=data;  
    }  
}
```

## 9. Clase Urls

```
public class Urls {  
    public int urlId;  
    public String url;  
  
    public Urls(int id, String data)  
    {  
        urlId=id;  
        url=data;  
    }  
}
```

## 10. Clase Texto



```

import java.io.BufferedWriter;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.OutputStreamWriter;
import java.io.Writer;
import java.util.ArrayList;

```

```

public class Texto {

```

```

    int posInicial;
    int posFinal;
    int posUrl;

```

```

    public Texto() {
        this.posInicial = 0;
        this.posFinal = 0;
        this.posUrl = 0;
    }

```

### 10.1. Obtener Texto

```

public String getTexto(String texto) {

    String textoReemplazarEtiqueta = "";
    String textoReemplazar = "";
    boolean _continue = true;

```

### 10.2. Reemplazar Signos

```

    texto = texto.replace("(", " ").replace(")", " ");
    texto = texto.replace("?", " ");
    texto = texto.replace(",", " ");
    texto = texto.replace("*", " ");
    texto = texto.replace("&", " ");
    texto = texto.replace("$", " ");
    texto = texto.replace(";", " ");
    texto = texto.replace("%", " ");
    texto = texto.replace("^", " ");
    texto = texto.replace("-", " ");
    texto = texto.replace("»", " ");
    texto = texto.replace("\\", " ");
    texto = texto.replace("!", " ");
    texto = texto.replace("|", " ");
    texto = texto.replace("+", " ");
    texto = texto.replace("[", " ");
    char texto1= "";
    char tex= ' ';
    texto = texto.replace(texto1,tex);
    texto = texto.replace("","");

```

### 10.3. Reemplazar Etiqueta y Menciones

```
while (_continue) {
    int start = texto.indexOf("#");
    if (start == -1) {
        start = texto.indexOf("@");
    }
    if (start != -1) {
        posInicial = texto.indexOf(":", start + 1);
        posFinal = texto.indexOf(" ", start + 1);
        if (posInicial != -1) {
            if (posInicial > posFinal) {
                posInicial = posFinal;
            } else {
                posFinal = posInicial;
            }
        }
        if (posFinal != -1) {
            if (posInicial != -1 && posFinal > posInicial) {
                posFinal = posInicial;
            } else {
                posInicial = posFinal;
            }
        }
        if (posInicial != -1) {
            textoReemplazar = texto.substring(start, posInicial);
        } else if (start != -1) {
            textoReemplazar = texto.substring(start, texto.length());
        }
        } else {
            _continue = false;
        }

        if (posInicial != -1) {
            textoReemplazarEtiqueta = texto.replaceFirst(textoReemplazar, "");
        } else if (start != -1) {
            textoReemplazarEtiqueta = texto.replaceFirst(textoReemplazar, "");
        } else {
            textoReemplazarEtiqueta = texto;
        }
        texto = textoReemplazarEtiqueta;
    }
}
```

## 10.4. Reemplazar Urls

```
_continue = true;
while (_continue) {
    int start = texto.indexOf("http://");
    if (start != -1) {
        posUrl = texto.indexOf(" ", start + 1);
        if (posUrl != -1) {
            textoReemplazar = texto.substring(start, posUrl);
        } else {
            textoReemplazar = texto.substring(start, texto.length());
        }
    } else {
        start = texto.indexOf("https://");
        if (start != -1) {
            posUrl = texto.indexOf(" ", start + 1);
            if (posUrl != -1) {
                textoReemplazar = texto.substring(start, posUrl);
            } else {
                textoReemplazar = texto.substring(start, texto.length());
            }
        }
    }
    if (start == -1) {
        _continue = false;
    } else {
        if (posUrl != -1) {
            textoReemplazar = texto.replaceFirst(textoReemplazar, "");
            texto = textoReemplazar;
        } else {
            textoReemplazar = texto.replaceFirst(textoReemplazar, "");
            texto = textoReemplazar;
        }
    }
    posUrl = -1;
}
```

## 10.5. Reemplazar otros Signos

```
String _texto= texto.replace(".", "").replace("|", "").replaceAll(":", "").replaceAll("-",
 "").replaceAll("/", "").replaceAll("#", ""); //.replaceAll("[", "").replaceAll("]", "").replaceAll("{",
 "").replaceAll("}", "");
return (_texto);
}
```

## 11. CLASE MainTwit

```
import Logica.Etiquetas;
import Logica.Menciones;
import Logica.Texto;
import Logica.Textos;
import Logica.Tipos;
import Logica.Twits_Query;
import Logica.Url;
import java.io.BufferedWriter;
import java.io.Console;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.OutputStreamWriter;
import java.io.Writer;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.util.ArrayList;
import java.util.Vector;
import java.util.logging.Level;
import java.util.logging.Logger;
import tweet.Cambios;

public class MainTwit extends javax.swing.JFrame {
    TableModel tableModel;
    public static final String[] columnNames = {
        "Url", "# tweet", "Etiqueta", "Mencion", "Url", "Texto", "Total",};

    /**
     * Creates new form NewJFrame
     */

    public MainTwit() {
        initComponents();
        tableModel = new TableModel(columnNames);
        jTable1.setModel(tableModel);
        jTable1.setSurrendersFocusOnKeystroke(true);
    }

    /**
     * This method is called from within the constructor to initialize the form.
     * WARNING: Do NOT modify this code. The content of this method is always
     * regenerated by the Form Editor.
     */

    @SuppressWarnings("unchecked")

    // <editor-fold defaultstate="collapsed" desc="Generated Code">
    private void initComponents() {
        jbtProcesar = new javax.swing.JButton();
        jScrollPane1 = new javax.swing.JScrollPane();
    }
}
```

```

jTable1 = new javax.swing.JTable();
jButton1 = new javax.swing.JButton();

        setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
jbtProcesar.setText("Procesar Twit");
jbtProcesar.setActionCommand("jbtProcesar");
jbtProcesar.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        jbtProcesarActionPerformed(evt);
    }
});

        jTable1.setModel(new javax.swing.table.DefaultTableModel(
            new Object [][] {

            },
            new String [] {

            }
        ));
jScrollPane1.setViewportView(jTable1);
jButton1.setText("Presentar Resultados");
jButton1.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        jButton1ActionPerformed(evt);
    }
});

javax.swing.GroupLayout layout = new javax.swing.GroupLayout(getContentPane());
getContentPane().setLayout(layout);
layout.setHorizontalGroup(
    layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addGroup(layout.createSequentialGroup()
            .addGap(18, 18, 18)
            .addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
                .addComponent(jScrollPane1, javax.swing.GroupLayout.DEFAULT_SIZE, 851, Short.MAX_VALUE)
                .addGroup(layout.createSequentialGroup()
                    .addComponent(jbtProcesar, javax.swing.GroupLayout.PREFERRED_SIZE, 258, javax.swing.GroupLayout.PREFERRED_SIZE)
                    .addGap(35, 35, 35)
                    .addComponent(jButton1, javax.swing.GroupLayout.PREFERRED_SIZE, 225, javax.swing.GroupLayout.PREFERRED_SIZE)
                    .addGap(0, 0, Short.MAX_VALUE))
                .addContainerGap())
            .addContainerGap());
layout.setVerticalGroup(
    layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
        .addGroup(layout.createSequentialGroup()
            .addGap(18, 18, 18)
            .addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
                .addComponent(jScrollPane1, javax.swing.GroupLayout.DEFAULT_SIZE, 851, Short.MAX_VALUE)
                .addGroup(layout.createSequentialGroup()
                    .addComponent(jbtProcesar, javax.swing.GroupLayout.PREFERRED_SIZE, 258, javax.swing.GroupLayout.PREFERRED_SIZE)
                    .addGap(35, 35, 35)
                    .addComponent(jButton1, javax.swing.GroupLayout.PREFERRED_SIZE, 225, javax.swing.GroupLayout.PREFERRED_SIZE)
                    .addGap(0, 0, Short.MAX_VALUE))
                .addContainerGap())
            .addContainerGap());

```

```

.addComponent(jbtProcesar)
.addComponent(jButton1))
.addGap(18, 18, 18)
.addComponent(jScrollPane1, javax.swing.GroupLayout.DEFAULT_SIZE, 498,
Short.MAX_VALUE)
.addContainerGap()
);

```

```

jbtProcesar.getAccessibleContext().setAccessibleName("jbtProcesar");
pack();
} // </editor-fold>
private void jbtProcesarActionPerformed(java.awt.event.ActionEvent evt) {
    // TODO add your handling code here:
    Twits_Query tw = new Twits_Query();
    ArrayList etiquetaList = new ArrayList();
    ArrayList mencionList = new ArrayList();
    ArrayList urlList = new ArrayList();
    ArrayList textoList = new ArrayList();

```

```

    try {
    ResultSet rs = tw.getUrlId();
    while (rs.next()) {

```

```

int urlId = Integer.parseInt(rs.getString("URL_ID"));
int num_tweet = Integer.parseInt(rs.getString("num"));

```

### 11.1. Añade las etiquetas, menciones, urls y texto en un arrayList

```

ResultSet rsTwit = tw.getTwits(urlId);
    while (rsTwit.next()) {
        int twit = Integer.parseInt(rsTwit.getString("TWETT_ID"));

        Etiquetas etiqueta = new Etiquetas(twit, tw.getEtiqueta(twit));
        etiquetaList.add(etiqueta);

        Menciones mencion = new Menciones(twit, tw.getMencion(twit));
        mencionList.add(mencion);

        Urls url = new Urls(twit, tw.getUrl(twit));
        urlList.add(url);

        Texto texto = new Texto();
        Textos textos = new Textos(twit, texto.getTexto(tw.getTexto(twit)));

```

```
textoList.add(textos);
}
```

## 11.2. Mide los cambios de las etiquetas, menciones urls y texto.

```
Tipos comparacion = new Tipos();
ArrayList<Cambios> repEtiqueta = comparacion.getNumEtiqueta(etiquetaList,
etiquetaList);
```

```
Tipos cambiosMenciones = new Tipos();
ArrayList<Cambios> repMenciones = cambiosMenciones.getNumMenciones(mencionList, mencionList);
```

```
Tipos cambiosUrls = new Tipos();
ArrayList<Cambios> repUrls = cambiosUrls.getNumUrls(urlList, urlList);
```

```
Tipos cambiosTextos = new Tipos();
ArrayList<Cambios> repTextos = cambiosTextos.getNumTextos(textoList,
textoList);
```

```
int ietiqueta=0;
int imencion=0;
int iurl=0;
int itexto=0;
if (repEtiqueta.size(>0)
{
    ietiqueta=repEtiqueta.size()-1;
}
if (repMenciones.size(>0)
{
    imencion=repMenciones.size()-1;
}
if (repUrls.size(>0)
{
    iurl=repUrls.size()-1;
}
```

```

        if (repTextos.size()>0)
        {
            itexto=repTextos.size()-1;
        }

```

### 11.3. Promedio de los cambios en las Métricas

```

            int total = (int) (ietiqueta+imencion+iurl+itexto) / 4;

            tw.ins_twit(urlId, num_tweet, imencion, ietiqueta, iurl,itexto, total);
        }
    } catch (Exception ex) {
        Logger.getLogger(MainTwit.class.getName()).log(Level.SEVERE, null, ex);
    }
}

```

```

private void jButton1ActionPerformed(java.awt.event.ActionEvent evt) {
    // TODO add your handling code here:
    Vector data = new Vector();
    Twits twis = new Twits();
    Twits_Query tw = new Twits_Query();
    ResultSet rs;
    try {
        rs = tw.getTwit();

        while (rs.next()) {
            twis.id = Integer.parseInt(rs.getString("gp_url_id"));
            twis.twit = Integer.parseInt(rs.getString("gp_num_tweet"));
            twis.mencion = Integer.parseInt(rs.getString("gp_mencion"));
            twis.etiqueta = Integer.parseInt(rs.getString("gp_etiqueta"));
            twis.url = Integer.parseInt(rs.getString("gp_url"));
            twis.texto = Integer.parseInt(rs.getString("gp_texto"));
            twis.total = (int) (twis.mencion + twis.etiqueta + twis.url + twis.texto) / 4;
            data.add(twis);
            twis = new Twits();
        }
        tableModel.agregarFila(data);
    }
}

```



```

    } catch (SQLException ex) {
        Logger.getLogger(MainTwit.class.getName()).log(Level.SEVERE, null, ex);
    }
}

/**
 * @param args the command line arguments
 */
public static void main(String args[]) {
    /* Set the Nimbus look and feel */
    //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code (optional) ">
    /* If Nimbus (introduced in Java SE 6) is not available, stay with the default look and feel.
    * For details see
    http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html
    */
    try {
        for (javax.swing.UIManager.LookAndFeelInfo info :
            javax.swing.UIManager.getInstalledLookAndFeels()) {
            if ("Nimbus".equals(info.getName())) {
                javax.swing.UIManager.setLookAndFeel(info.getClassName());
                break;
            }
        }
    } catch (ClassNotFoundException ex) {

        java.util.logging.Logger.getLogger(MainTwit.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (InstantiationException ex) {

        java.util.logging.Logger.getLogger(MainTwit.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (IllegalAccessException ex) {

        java.util.logging.Logger.getLogger(MainTwit.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (javax.swing.UnsupportedLookAndFeelException ex) {

```

```

java.util.logging.Logger.getLogger(MainTwit.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    }
//</editor-fold>
/* Create and display the form */
java.awt.EventQueue.invokeLater(new Runnable() {
    public void run() {
        new MainTwit().setVisible(true);
    }
});
}
// Variables declaration - do not modify
private javax.swing.JButton jButton1;
private javax.swing.JScrollPane jScrollPane1;
private javax.swing.JTable jTable1;
private javax.swing.JButton jButtonProcesar;
// End of variables declaration
}

```

## 12. CLASE TABLE MODEL

```

import java.util.Collection;
import java.util.Vector;
import javax.swing.table.AbstractTableModel;

public class TableModel extends AbstractTableModel {
    public static final int ID_INDEX = 0;
    public static final int TWIT_INDEX = 1;
    public static final int ETIQUETA_INDEX = 2;
    public static final int MENCION_INDEX = 3;
    public static final int URL_INDEX = 4;
    public static final int TEXTO_INDEX = 5;
    public static final int TOTAL_INDEX = 6;

    protected String[] columnasNombre;
    protected Vector data;

    public TableModel(String[] columnas) {
        this.columnasNombre = columnas;
        data = new Vector();
    }
}

```

```

public String getColumnName(int column) {
    return columnasNombre[column];
}

public boolean isCellEditable(int row, int column) {
    return false;
}

public Class getColumnClass(int column) {
    switch (column) {
    case ID_INDEX:
    case TWIT_INDEX :
    case ETIQUETA_INDEX:
    case MENCION_INDEX :
    case URL_INDEX :
    case TEXTO_INDEX:
    case TOTAL_INDEX:
    default:
    return Object.class;
    }
}

public Object getValueAt(int row, int column) {
    Twits record = (Twits)data.get(row);
    switch (column) {
    case ID_INDEX :
    return record.getId();
    case TWIT_INDEX :
    return record.getTwit();
    case ETIQUETA_INDEX :
    return record.getEtiqueta();
    case MENCION_INDEX :
    return record.getMencion();
    case URL_INDEX :
    return record.getUrl();
    case TEXTO_INDEX :
    return record.getTexto();
    case TOTAL_INDEX :
    return record.getTotal();
    default:
    return new Object();
    }
}

public void setValueAt(Object value, int row, int column) {
    Twits record = (Twits)data.get(row);
    switch (column) {
    case ID_INDEX :
    record.setId((int)value);
    case TWIT_INDEX :

```

```

        record.setTwit((int)value);
        case ETIQUETA_INDEX :
        record.setEtiqueta((int)value);
        case MENCION_INDEX :
        record.setMencion((int)value);
        case URL_INDEX :
        record.setUrl((int)value);
        case TEXTO_INDEX :
        record.setTexto((int)value);
        case TOTAL_INDEX :
        record.setTotal((int)value);
        default:
        System.out.println("");
    }
    fireTableCellUpdated(row, column);
}

public int getRowCount() {
    return data.size();
}

public int getColumnCount() {
    return columnasNombre.length;
}

public boolean hasEmptyRow() {
    if (data.size() == 0) return false;
    Twits record = (Twits)data.get(data.size() - 1);
    if (record.getEtiqueta()==0 &&
        record.getMencion()==0 &&
        record.getTexto()==0&&
        record.getTwit()==0&&
        record.getUrl()==0)
    {
        return true;
    }
    else return false;
}

public void agregarFila(Vector tw) {
    data=tw;
    fireTableRowsInserted(
        data.size() - 1,
        data.size() - 1);
}

```

### 13. CLASE Twits

```

public class Twits {
    protected int mencion;
    protected int url;
}

```

```

protected int texto;
protected int etiqueta;
protected int twit;
protected int total;
protected int id;

public Twits() {
    mencion = 0;
    url = 0;
    texto = 0;
    etiqueta=0;
    twit = 0;
    id = 0;
}

public int getId() {
    return id;
}

public void setId(int id) {
    this.id = id;
}

public int getTwit() {
    return twit;
}

public void setTwit(int twit) {
    this.twit = twit;
}

public int getMencion() {
    return mencion;
}

public void setMencion(int mencion) {
    this.mencion = mencion;
}

public int getUrl() {
    return url;
}

public void setUrl(int url) {
    this.url = url;
}

public int getTexto() {
    return texto;
}

```

```

public void setTexto(int texto) {
    this.texto = texto;
}

public int getEtiqueta() {
    return etiqueta;
}

public void setEtiqueta(int etiqueta) {
    this.etiqueta = etiqueta;
}

public int getTotal() {
    return total;
}

public void setTotal(int total) {
    this.total = total;
}
}

```

## 15. URLs para descargar los libros de actas de los papers.

### ÁTICA 2013

V Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2013), Proyecto ESVI-AL, Universidad Continental, Huancayo, (Perú, 2 - 4 de octubre 2013).

<b>Estado del Arte sobre Estudio de los Cambios en la Descripción de Recursos Publicados en las Redes Sociales. Aplicación en Twitter</b>
---

Libro de actas <a href="http://www.esvial.org/atica2013/documentos/LibroATICA2013.pdf">http://www.esvial.org/atica2013/documentos/LibroATICA2013.pdf</a>
--

### CAFVIR 2014

V Congreso Internacional sobre Calidad y Accesibilidad de la Formación Virtual (CAFVIR 2014), Proyecto ESVI-AL, Universidad Galileo, Antigua Guatemala (Guatemala, 14-16 de Mayo de 2014).

<b>Cambios en la Descripción de Recursos Publicados en las Redes Sociales. Aplicación en Twitter</b>
--

Libro de actas <a href="http://www.esvial.org/cafvir2014/documentos/LibroActasCAFVIR2014.pdf">http://www.esvial.org/cafvir2014/documentos/LibroActasCAFVIR2014.pdf</a>
--

# Estado del Arte sobre Estudio de los Cambios en la Descripción de Recursos Publicados en las Redes Sociales Aplicación en Twitter

Gabriela M. Agila<sup>1</sup>, y Jorge A. López<sup>1</sup>

<sup>1</sup>Tecnologías Avanzadas de la Web y Sistemas Basados en el Conocimiento  
Titulación en Sistemas Informáticos y Computación.  
Universidad Técnica Particular de Loja  
1101608 San Cayetano Alto S/N (Loja-Ecuador)  
E-mail: {gmagila, jalopez2}@utpl.edu.ec

**Resumen.** En este artículo se presenta una investigación en desarrollo sobre los cambios en la descripción de recursos publicados como links en la red social Twitter, los cambios serán medidos a través de hashtags, menciones, y palabras. Con este estudio se pretende diseñar e implementar un algoritmo que determine los cambios que experimenta un recurso, para así poder realizar un estudio real de la difusión de información en Twitter.

**Palabras clave:** Hashtags, Menciones, Palabras, SNA, Difusión de la Información.

## 1 Introducción

Las Redes Sociales son herramientas útiles que permiten informar, comunicar y conectar con personas en cualquier parte del mundo. El acceso al Internet es diferente en cada región, sin embargo existe una necesidad latente del ser humano por interactuar con otros que han hecho que esas brechas digitales sean cada vez menores (iLifebelt, 2013).

En sentido amplio, una red social es una estructura social formada por personas o entidades conectadas y unidas entre sí por algún tipo de relación o interés común. El término se atribuye a los antropólogos británicos Alfred Radcliffe-Brown y Jhon Barnes. Las redes sociales son parte de nuestra vida, son la forma en la que se

estructuran las relaciones personales, estando conectados mucho antes de tener conexión a Internet (Ponce, 2012).

Los medios de comunicación revolucionaron la vida de las personas a medida que fueron apareciendo en su contexto de emergencia. Esto ocurrió en nuestro país y en el mundo entero. Internet no ha sido la excepción. Desde su surgimiento cambió rotundamente la forma de comunicación que se manejaba hasta entonces. Fue evolucionando a pasos agigantados hasta convertirse en lo que hoy conocemos. Las redes sociales, como la que se tratará en esta investigación, Twitter, que es una manera nueva de comunicarse y de reencontrarse con personas activas en esta red.

“El éxito de Twitter se basa en la simplicidad de su idea matriz”, (Orihuela) . Twitter se promociona como una de las mejores maneras de compartir y descubrir qué está pasando a nuestro alrededor en este preciso momento sobre todo para acceder a la vida de personalidades importantes , se trata de comprender qué usos se le da hoy a esta red, sobre todo en la comunicación y la transmisión de información.

Es por esto que se realiza una investigación para estudiar los cambios en la descripción de recursos que contengan links publicados en la red social, específicamente Twitter en donde se determinará el cambio del “tweet” con respecto a los “hashtags”, “menciones”, y “palabras”. Para este estudio se presentan métricas y medidas que se analizarán para diseñar e implementar un algoritmo que determine los cambios de los recursos.

Este documento está estructurado de la siguiente manera: una sección donde se describe el análisis de redes sociales y sus métricas. Una segunda sección que habla sobre la difusión de la información. Como tercera sección se presenta trabajos relacionados. Una cuarta donde se presenta la propuesta de análisis de difusión de información. Finalmente una sección de los resultados que se pretende obtener.

## **2 Difusión de la Información**

La difusión de la información es poder sólo cuando la información es compartida y produce conocimiento colectivo susceptible de mejorar cualquier aspecto de la sociedad. En este sentido las redes sociales están jugando un papel revolucionario pero dónde perderse en el anonimato es cada vez más fácil. (Orta, Pardo, & Salas). Potencialmente, la difusión de información en las redes sociales puede ser muy eficiente y la información puede llegar a un gran número de personas en un período muy corto de tiempo. Sin embargo, poco se sabe acerca de los patrones de difusión de información en los medios sociales y cómo el comportamiento de información en línea puede haber cambiado.

La influencia se traduce en la capacidad de hacer llegar el mensaje al mayor número de individuos posibles, donde la información en twitter fluye de manera viral a partir de la creación y generación de un mensaje y su posterior redifusión por la red de influencia de la persona que lo emite, y de las redes de los que lo reciben y deciden reemitirlo, y así sucesivamente hasta que el mensaje muere (Orta, Pardo, & Salas). Asimismo se puede definir “influencia” como la probabilidad asociada a un nodo de transmitir o impedir la transmisión de nuevas ideas o pautas de comportamiento en la red (Ugarte).



Los expertos coinciden que la aparición de Internet y los navegadores es un hito en el ámbito de la difusión de información. Por ejemplo, la web 2.0 ha jugado un papel muy importante en el desarrollo de sistemas descentralizados de distribución de información. Sistemas como las páginas wiki (Wikipedia), las bitácoras o microblogs (Twitter), los sistemas de alertas (RSS) o los sistemas de votación (Menéame) han permitido la construcción de comunidades virtuales que aprovechan la inteligencia colectiva para generar conocimiento y difundirlo (De la Rosa Troyano).

Twitter tiene una amplia colaboración para la difusión de información, mediante sus servicios, la difusión del hashtag como tema está gobernada por la influencia de aquellos que lo utilizan (Romero, Meeder, & Kleinberg, 2011). Esto es una gran ventaja porque ayuda a organizar la información para que pueda ser difundida y buscada de mejor manera.

Cada día millones de usuarios se conectan a través de redes sociales, generando un rico tesoro de datos que permiten estudiar los mecanismos que subyacen a las interacciones humanas, la estructura de la red afecta a la propagación de la información entre los usuarios, la red es a su vez formada por esta actividad de comunicación. Pero los accesos directos basados en el tráfico son otro factor clave en la interpretación de la evolución de la red. Los usuarios que son populares, activos e influyentes tienden a crear accesos directos por el tráfico, haciendo que el proceso de difusión de la información sea más eficiente en la red (Weng, y otros, 2013).

## 2.1 Twitter

Es una red de información en tiempo real, siendo un servicio de microblogging más popular y extendido en la Red, en Twitter existen pequeñas explosiones de información llamadas Tweets (mensajes) que se publican de no más de 140 caracteres de longitud, que te conectan con las últimas historias, ideas, opiniones y noticias sobre lo que encuentras interesante como: cine, gobierno y política, gastronomía, organizaciones, entretenimiento, fútbol, música, cuentas oficiales, elecciones, innovación, turismo, deportes, medios y periodistas, literatura, moda y belleza. Para esto se debe localizar las cuentas que te resulten más atractivas y podrás seguir sus conversaciones, convirtiéndose así en un punto de encuentro global (Twitter, 2013).

La red social twitter se compone de followers y following, que no tienen por qué coincidir. Intrínsecamente estas redes se verán conformadas por usuarios de intereses similares, bien sean políticos, económicos, de ocio, etc., lo que se denomina temas o grupos semánticos primarios (Orta, Pardo, & Salas).

## 2.2 Datos dentro del contenido del tweet

La información que se encuentra dentro del contenido del tweet, se la menciona a continuación:

**Etiquetas / Hashtags.-** Es una convención adoptada por los usuarios de Twitter para relacionar sus mensajes y hacer referencia a un tópico o tema específico. (Carter, Tsagkias, & Weerkamp, 2011). El hashtag se compone de una cadena de caracteres

corta precedida del símbolo #, fue un recurso creado orgánicamente por la comunidad de usuarios de Twitter, como una manera de categorizar los mensajes.

Los hashtags reúnen información relativa a un tema y su divulgación y dispersión depende de la red de los seguidores que deseen participar en la conversación. La difusión del hashtag como tema está gobernada por la influencia de aquellos que lo utilizan (Romero, Meeder, & Kleinberg, 2011). La ventaja es que ayudan a organizar la información para que pueda ser difundida y buscada de mejor manera.

**Menciones.-** Una mención es cualquier actualización de Twitter que contenga “@nombredeusuario” en el cuerpo del Tweet, sin importar en qué parte del mensaje está, las @respuestas también son consideradas menciones (Twitter, 2013).

Las menciones, no siempre le hablan al usuario sino que en ocasiones se usa para referir a un usuario, claro buscando que la persona tenga conocimiento de esto. Los tweets que incluyen menciones son empleados para informar al destinatario que se ha publicado un mensaje que está dirigido a él, se trata de un mensaje, en principio público y visible por cualquiera, que en el perfil del usuario destinatario irá marcado de forma especial para darle prioridad. (Collado C, 2011).

**Palabras.-** Es útil que el tweet contenga texto, que es un entramado de signos con una intención comunicativa que busca transmitir un cierto mensaje que adquiere sentido de acuerdo a cada contexto (Wikipedia), pudiendo ser una idea, opinión o conversación acerca de algún asunto o tema en específico. Se puede aislar las menciones, hashtag y url, para tomar solo el texto y poder aplicar algún método que permita analizar palabras.

**Links.-** Las URL son un Localizador Uniforme de Recursos siendo un sitio web que apunta a una página única en la Internet (Twitter, 2013) . Se ha demostrado que los tweets más retweets son los que comparten enlaces o links en su contenido. Las personas se interesan más por tweets que contienen una url con información adicional.

**URL Des-acortada.-** El acortamiento de URL es una técnica en la World Wide Web. Esto se logra mediante el uso de una redirección HTTP en un nombre de dominio que sea corto, que conecta a la página web que tiene una larga URL. En el que un localizador uniforme de recursos (URL) podrá hacerse sustancialmente más corto en longitud y sigue directamente a la página deseada (Burply). Esto es especialmente conveniente para la tecnología de mensajería como Twitter que limita severamente el número de caracteres que pueden utilizarse en un mensaje.

## 2.3 Indicadores Twitter

Las redes sociales como Twitter parten principalmente de los motores de búsqueda, pero arrastran influencias de los medios de comunicación tradicionales y así se habla de ciertos indicadores, estándares de facto en la actualidad, (Orta, Pardo, & Salas) como:

- Seguidores: Es el número total de seguidores del usuario.
- Influencia de red: No sólo tiene en cuenta tus seguidores sino su influencia.

- Amplificación de red: Valor que muestra la influencia directa que tiene un usuario sobre su red.
- Alcance real: Es un indicador que a partir de la influencia de red y la amplificación de red calcula el valor de audiencia real a la que puede llegar el usuario con un mensaje.
- Otras medidas relacionadas con la pasión, la fuerza, el alcance de los mensajes emitidos por un usuario.

En todos estos casos se parte de un análisis de red social creada a partir de los seguidores y los seguidos, y de los mensajes que se reenvían de un usuario, pero todos ellos son análisis centrados en la influencia global de un usuario y no tienen en cuenta el resto de métricas que el ARS puede extraer de una determinada red, y utilizar estas métricas para sugerir influenciadores a partir de un determinado tema (Orta, Pardo, & Salas).

Estas tres actividades representan los diferentes tipos de influencia de una persona: (Cha, Haddadi, Benevenuto, & Gummadi).

1. **Influencia grado de entrada**, el número de seguidores de un usuario, indica directamente el tamaño de la audiencia para ese usuario.

2. **Influencia Retweet**, que medimos a través del número de retweets que contengan el nombre de uno, indica la capacidad de ese usuario para generar contenido con paso de largo el valor.

3. **Mención de influencia**, que se mide a través del número de menciones que contiene el nombre de uno, indica la capacidad de dicho usuario para involucrar a otros en una conversación.

### 3 Trabajos Relacionados

#### **Contagio de la influencia basada en la difusión de homofilia impulsada en redes dinámicas.**

La difusión depende del tipo de contagio. Contagios complejos, a diferencia de las enfermedades infecciosas (contagios simples), se ven afectados por el refuerzo social y homofilia. Por lo tanto, la propagación en comunidades altamente clúster es mayor, mientras que la difusión a través de las comunidades se ve obstaculizada. Una hipótesis común es que los memes y comportamientos son contagios complejos. Se demuestra que, si bien la mayoría de los memes de hecho se extendió como contagios complejos, algunos memes virales se propagan a través de muchas comunidades, como las enfermedades. Las características de nodo y comportamientos a menudo se relacionan con la estructura de las redes sociales a través del tiempo. (Weng, Menczer, & Ahn, 2013).

#### **Diferencias en la mecánica de difusión de la información.**

El trabajo trata sobre la “Variación en la propagación de hashtags”. Donde se analizan las fuentes de variación en la forma de que los hashtags más utilizados en Twitter se propagan dentro de su población de usuarios. Se Encuentra que estas fuentes de variación no sólo implican diferencias en la probabilidad con la que algo se contagia de una persona a otra. (Romero, Meeder, & Kleinberg, 2011)

### **Impacto de los patrones de actividad humanas sobre la dinámica de difusión de la información.**

Estudia el impacto de los patrones de la actividad humana sobre la difusión de información. Con un experimento de correo electrónico viral que incluyó 31.183 personas en el que han sido capaces de rastrear una pieza específica de información a través de la red social. Se ha encontrado que la información viaja a un ritmo lento inesperadamente. Mediante el uso de un modelo de ramificación que describe con precisión el experimento, se muestra que la gran heterogeneidad que se encuentra en el tiempo de respuesta es el responsable de la lenta dinámica de la información a nivel colectivo. (Iribarren, 2009).

### **Propuesta**

Potencialmente, la difusión de información en las redes sociales puede ser muy eficiente y la información puede llegar a un gran número de personas en un período muy corto de tiempo. Sin embargo, poco se sabe acerca de los patrones de difusión de información en los medios sociales y cómo el comportamiento de información en línea puede haber cambiado.

Es por esto que en la presente investigación en desarrollo se propone estudiar los cambios en la descripción de recursos “Links” publicados en Twitter, analizando la difusión de la información, en esta red social, a través del cambio en hashtags, menciones, y palabras. Con este estudio se pretende diseñar e implementar un algoritmo que determine los cambios, para así poder realizar un estudio real de los mismos.

### **4 Resultados Esperados**

Los resultados que se esperan obtener son el diseño y la implementación de un algoritmo que permita determinar los cambios en la descripción de los recursos “links” de twitter, consiguiendo resultados que faciliten el análisis de la difusión de información en esta red social. Para culminar con un estudio real de los cambios que se han dado.

### **Bibliografía**

*Twitter*. (2013). Obtenido de <http://www.twitter.com/>

*Burply*. (s.f.). Obtenido de <http://es.burply.com/hipertexto/url-acortamiento/tinyurl-312998.html>

- Carter, S., Tsagkias, M., & Weerkamp, W. (2011). *Twitter hashtags: Joint Translation and Clustering*. Obtenido de [http://www.webscil1.org/fileadmin/websci/Posters/125\\_paper.pdf](http://www.webscil1.org/fileadmin/websci/Posters/125_paper.pdf)
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (s.f.). *Measuring User Influence in Twitter: The Million Follower Fallacy*. Obtenido de Max Planck Institute for Software Systems (MPI-SWS), Germany; Royal Veterinary College, University of London, United Kingdom; CS Dept., Federal University of Minas Gerais (UFMG), Brazil: <http://www.google.com.ec/url?sa=t&rct=j&q=measuring%20user%20influence%20in%20twitter%3A%20the%20million%20follower%20fallacy&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.aaai.org%2Focs%2Findex.php%2FICWSM%2FICWSM10%2Fpaper%2Fdownload%2F1538%2F1826&>
- Collado C, E. (2011). *Máster universitario de Sociedad de la Información y el Conocimiento*. Obtenido de <http://eduangi.com/wp-content/uploads/2011/07/El-microblogging-en-el-periodismo-esp%C3%B1ol.pdf>
- De la Rosa Troyano, F. (. (s.f.). *Sistemas de Inteligencia Web: Análisis de Redes Sociales*. Obtenido de <http://www.lsi.us.es/~ffrosat/publicaciones/Tesis.2011.v75.vf.pdf>
- iLifebelt. (2013). *Estudio Redes Sociales Centroamérica 2013*. Obtenido de II Estudio de las Redes Sociales en América Central: <http://octavioislas.files.wordpress.com/2013/08/redes-sociales-centroamerica-2013.pdf>
- Iribarren, J. (17 de Julio de 2009). *Impact of Human Activity Patterns on the Dynamics of Information Diffusion*. Obtenido de IBM Corporation, ibm.com e-Relationship Marketing Europe, 28002 Madrid, Spain: <http://prl.aps.org/abstract/PRL/v103/i3/e038702>
- Orihuela, J. (s.f.). *Profesor en la Facultad de Comunicación de la Universidad de Navarra y autor del blog eCuaderno*. Obtenido de <http://www.virtualogos.net/portal/virtual-community/virtual-teens/pliteraria/147-twitter-como-medio-de-comunicacion.html>
- Orta, C., Pardo, F., & Salas, J. (s.f.). *Análisis de Redes de Influencia en Twitter*. Obtenido de <http://www.google.com.ec/url?sa=t&rct=j&q=an%C3%A1lisis%20de%20redes%20de%20influencia%20en%20twitter&source=web&cd=1&ved=0CDsQFjAA&url=http%3A%2F%2Fwww.redesymediossociales.es%2Fdescarga.php%3Ffile%3DAnalisis%2520de%2520influencia%2520en%2520Twitter.pdf&>
- Ponce, I. (17 de Abril de 2012). (17/4/2012): " *Monográfico: Redes Sociales*". *Observatorio Tecnológico NIPO:820-10-289-9*. Obtenido de <http://recursostic.educacion.es/observatorio/web/es/internet/web-20/1043-redes-sociales>
- Romero, D., Meeder, B., & Kleinberg, J. (Junio de 2011). " *Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter*". Obtenido de (ed.)

- International World Wide Web Conference Committee (IW3C2).:  
<http://www.cs.cornell.edu/home/kleinber/www11-hashtags.pdf>
- Twitter. (2013). *El glosario de Twitter*. Obtenido de  
<https://support.twitter.com/articles/352810-el-glosario-de-twitter#>
- Ugarte, D. d. (s.f.). *Breve Historia del análisis de redes sociales*. Obtenido de  
[http://lasindias.com/gomi/historia\\_del\\_analisis\\_de\\_redes\\_sociales.pdf](http://lasindias.com/gomi/historia_del_analisis_de_redes_sociales.pdf)
- Weng, L., Menczer, F., & Ahn, Y.-Y. (28 de August de 2013). *Virality Prediction and Community Structure in Social Networks*. Obtenido de Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA.:  
<http://www.nature.com/srep/2013/130828/srep02522/pdf/srep02522.pdf>
- Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., y otros. (20 de Junio de 2013). *The Role of Information Diffusion*. Obtenido de  
<http://arxiv.org/pdf/1302.6276v2.pdf>
- Wikipedia. (s.f.). *Wikipedia la enciclopedia libre*. Obtenido de Definición de texto:  
<http://es.wikipedia.org/wiki/Texto>

# Estudio de los Cambios en la Descripción de Recursos Publicados en las Redes Sociales Aplicación en Twitter

Gabriela M. Agila<sup>1</sup>, y Jorge A. López<sup>1</sup>

<sup>1</sup>Tecnologías Avanzadas de la Web y Sistemas Basados en el Conocimiento  
Titulación en Sistemas Informáticos y Computación.  
Universidad Técnica Particular de Loja  
1101608 San Cayetano Alto S/N (Loja-Ecuador)  
E-mail: {gmagila, jalopez2}@utpl.edu.ec

**Resumen.** En este artículo se presenta una investigación sobre los cambios en la descripción de recursos publicados como links en la red social Twitter, los que serán medidos a través de métricas como: hashtags, menciones, urls cortas y palabras. Con este estudio se diseñó e implementó un algoritmo que mide los cambios que experimenta un recurso, para así poder realizar un análisis real de la difusión de información en Twitter.

**Palabras clave:** Hashtags, Menciones, Palabras, SNA, Difusión de la Información, Twitter.

## 1 Introducción

Las Redes Sociales son herramientas útiles que permiten informar, comunicar y conectar con personas en cualquier parte del mundo. El acceso al Internet es diferente en cada región, sin embargo existe una necesidad latente del ser humano por interactuar con otros que han hecho que esas brechas digitales sean cada vez menores (iLifebelt, 2013).

Una red social es una estructura social formada por personas o entidades conectadas y unidas entre sí por algún tipo de relación o interés común. Las redes sociales son parte de nuestra vida, son la forma en la que se estructuran las relaciones personales, estando conectados mucho antes de tener conexión a Internet (Ponce, 2012).

“El éxito de Twitter se basa en la simplicidad de su idea matriz”, (Orihuela) . Twitter se promociona como una de las mejores maneras de compartir y descubrir qué está pasando a nuestro alrededor en este preciso momento, se trata de comprender qué usos se le da hoy a esta red, sobre todo en la comunicación y la transmisión de información.

Es por esto que se realizó una investigación para estudiar los cambios en la descripción de recursos que contengan links publicados en Twitter en donde se



determinará el cambio del “tweet” con respecto a los “hashtags”, “menciones”, “urls cortas” y “palabras”.

Este documento está estructurado de la siguiente manera: primera sección describe el análisis de redes sociales. Segunda sección habla sobre la difusión de la información. Como tercera sección una descripción de Twitter. Una cuarta sección que indica el diseño y la implementación del algoritmo que mide los cambios. En la quinta sección se presenta los resultados, mientras que en la sexta se encuentran las conclusiones. Finalmente una sección donde se presenta trabajos futuros.

## **2 Difusión de la Información**

La difusión de la información es poder sólo cuando la información es compartida y produce conocimiento colectivo susceptible de mejorar cualquier aspecto de la sociedad. En este sentido las redes sociales están jugando un papel revolucionario pero dónde perderse en el anonimato es cada vez más fácil. (Orta, Pardo, & Salas). Potencialmente, la difusión de información en las redes sociales puede ser muy eficiente y la información puede llegar a un gran número de personas en un período muy corto de tiempo (Iribarren, 2009). Sin embargo, poco se sabe acerca de los patrones de difusión de información en los medios sociales (Lilian Weng, Filippo Menczer , & Yong-Yeol Ahn, 2013) y cómo el comportamiento de información en línea puede haber cambiado.

La influencia se traduce en la capacidad de hacer llegar el mensaje al mayor número de individuos posibles, donde la información en twitter fluye de manera viral a partir de la creación y generación de un mensaje y su posterior redifusión por la red de influencia de la persona que lo emite, y de las redes de los que lo reciben y deciden reemitirlo, y así sucesivamente hasta que el mensaje muere (Orta, Pardo, & Salas). También se define “influencia” como la probabilidad asociada a un nodo de transmitir o impedir la transmisión de nuevas ideas o pautas de comportamiento en la red (Ugarte).

Los expertos coinciden que la aparición de Internet y los navegadores son un hito en el ámbito de la difusión de información. Por ejemplo, la web 2.0 ha contribuido en el desarrollo de sistemas descentralizados de distribución de información, como las páginas wiki (Wikipedia), las bitácoras o microblogs (Twitter), los sistemas de alertas (RSS) o los sistemas de votación (Menéame) han permitido la construcción de comunidades virtuales que aprovechan la inteligencia colectiva para generar conocimiento y difundirlo (De la Rosa Troyano F. ).

Twitter tiene una amplia colaboración para la difusión de información, mediante sus servicios, la difusión del hashtag como tema está gobernada por la influencia de aquellos que lo utilizan (Romero, Meeder, & Kleinberg, 2011). Esto es una gran ventaja porque ayuda a organizar la información para que pueda ser difundida y buscada de mejor manera.

Cada día millones de usuarios se conectan a través de redes sociales, generando un rico tesoro de datos que permiten estudiar los mecanismos que subyacen a las interacciones humanas, la estructura de la red afecta a la propagación de la información entre los usuarios, la red es a su vez formada por esta actividad de comunicación (Weng, y otros, 2013).



### 3 Twitter

Es una red de información en tiempo real, siendo un servicio de microblogging más popular y extendido en la Red, en Twitter existen pequeñas explosiones de información llamadas Tweets (mensajes) que se publican de no más de 140 caracteres de longitud, que te conectan con las últimas historias, ideas, opiniones y noticias sobre lo que encuentras interesante como: cine, gobierno y política, gastronomía, organizaciones, entretenimiento, fútbol, música, cuentas oficiales, elecciones, innovación, turismo, deportes, medios y periodistas, literatura, moda y belleza. Para esto se debe localizar las cuentas que te resulten más atractivas y podrás seguir sus conversaciones, convirtiéndose así en un punto de encuentro global (Twitter, 2013).

La red social twitter se compone de followers y following, que no tienen por qué coincidir. Intrínsecamente estas redes se verán conformadas por usuarios de intereses similares, bien sean políticos, económicos, de ocio, etc., lo que se denomina temas o grupos semánticos primarios (Orta, Pardo, & Salas).

#### 3.1 Datos dentro del contenido del tweet

La información que se encuentra dentro del contenido del tweet, se la menciona a continuación:

**Etiquetas / Hashtags.**- Es una convención adoptada por los usuarios de Twitter para relacionar sus mensajes y hacer referencia a un tópico o tema específico. (Carter, Tsagkias, & Weerkamp, 2011). El hashtag se compone de una cadena de caracteres corta precedida del símbolo #, fue un recurso creado orgánicamente por la comunidad de usuarios de Twitter, como una manera de categorizar los mensajes.

**Menciones.**- Una mención es cualquier actualización de Twitter que contenga “@nombredeusuario” en el cuerpo del Tweet, sin importar en qué parte del mensaje está, las @respuestas también son consideradas menciones (Twitter, 2013).

Las menciones, no siempre le hablan al usuario sino que en ocasiones se usa para referir a un usuario, claro buscando que la persona tenga conocimiento de esto (Collado C, 2011). Los tweets que incluyen menciones son empleados para informar al destinatario que se ha publicado un mensaje que está dirigido a él.

**Palabras.**- Es útil que el tweet contenga texto, que es un entramado de signos con una intención comunicativa que busca transmitir un cierto mensaje que adquiere sentido de acuerdo a cada contexto, pudiendo ser una idea, opinión o conversación acerca de algún asunto o tema en específico.

**Links.**- Las URL son un Localizador Uniforme de Recursos siendo un sitio web que apunta a una página única en la Internet (Twitter, 2013) . Se ha demostrado que los tweets más retwitteados son los que comparten enlaces o links en su contenido. Las personas se interesan más por tweets que contienen una url con información adicional.

**URL acortada.**- Es una técnica en la World Wide Web, esto se logra mediante el uso de una redirección HTTP en un nombre de dominio que sea corto, que conecta a la página web que tiene una larga URL. En el que una URL podrá hacerse sustancialmente más corto en longitud y sigue directamente a la página deseada. Esto

es conveniente para la tecnología de mensajería como Twitter que limita severamente el número de caracteres que pueden utilizarse en un mensaje.

### 3.2 Indicadores Twitter

Las redes sociales como Twitter parten principalmente de los motores de búsqueda, pero arrastran influencias de los medios de comunicación tradicionales y así se habla de ciertos indicadores (Orta, Pardo, & Salas), como:

- Seguidores: Es el número total de seguidores del usuario, lo que da una estimación de los que directamente visualizarán el mensaje que un usuario emita y por lo tanto puede ser una primera medida directa de su influencia, aunque no toma en cuenta más características de la subred del usuario.
- Influencia de red: No sólo tiene en cuenta tus seguidores sino su influencia, de manera recursiva se calcula el número de seguidores que tienen y la influencia de estos.
- Amplificación de red: Valor que muestra la influencia directa que tiene un usuario sobre su red, es decir, cuando emite un mensaje, cuál es la probabilidad de que ese mensaje sea retransmitido y por lo tanto amplificado.
- Alcance real: Es un indicador que a partir de la influencia de red y la amplificación de red calcula el valor de audiencia real a la que puede llegar el usuario con un mensaje.

Estas tres actividades representan los diferentes tipos de influencia de una persona (Cha, Haddadi, Benevenuto, & Gummadi):

1. **Influencia grado de entrada**, es el número de seguidores de un usuario que indica directamente el tamaño de su audiencia.

2. **Influencia Retweet**, se mide a través del número de retweets que contengan el nombre de uno, indica la capacidad de ese usuario para generar contenido con paso de largo el valor.

3. **Mención de influencia**, se mide a través del número de menciones que contiene el nombre de uno, indica la capacidad que tiene dicho usuario para involucrar a otros en una conversación.

### 3.3 Estructura del Tweet en twitter

La estructura indica el conjunto de todos los elementos importantes que forman un mensaje, se presentan básicamente dos estructuras de tweets (Jiménez , 2013), que se las ha categorizado de la siguiente manera:

**Estructura Simple:** Consiste en un tweet formado por un texto de 140 caracteres o menos.

**Estructura Compuesta:** A diferencia de los tweets de estructura simple, los de estructura compuesta se forman porque además del texto incluyen #etiquetas, @menciones, /via@nombreUsuario retwitteado, enlace acortado, obteniendo un tweet con sus todos sus elementos. No afecta si tiene una o más etiquetas, menciones, urls, texto, siempre y cuando no supere el límite de los caracteres que propicia la red social Twitter.

A continuación se muestran las convenciones en la estructura de un tweet, los elementos que lo conforman pueden ir en cualquier parte del mensaje:

Tweet para interactuar, compartir enlaces, y etiquetas:

- @Nombre del usuario + texto: Interactúa con otro usuario.
- Texto + enlace: Comparte enlaces.
- Texto + #hashtag: Comparte etiquetas.
- @Nombre del usuario + texto + #hashtag+ enlace: Interactúa, comparte enlaces y etiquetas.

Para hacer RT a un tweet:

- RT + @nombre de usuario retwitteado + Tweet
- Tweet + / vía @nombre del usuario retwitteado

## 4 Diseño e Implementación

A continuación se ha establecido las fases que se encuentran debidamente ordenadas para poder determinar el proceso que mida los cambios de las métricas planteadas, cuyos resultados sirvan finalmente para realizar un estudio real de los cambios obtenidos.

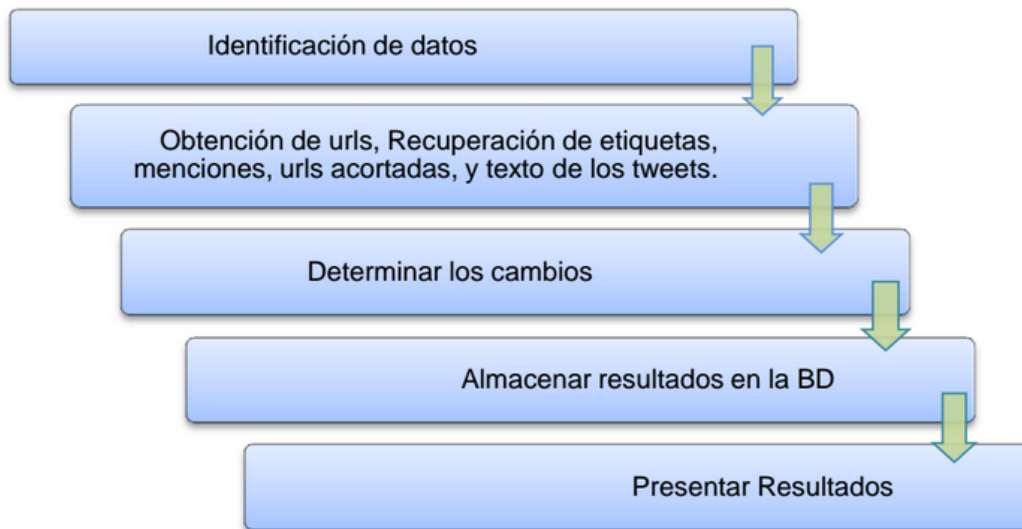


Figura 1: Proceso General para Determinar los Cambios

Se trabajó con NetBeans IDE 7.2.1 que es un entorno de desarrollo integrado, el cual ofrece soporte y es diseñado para el desarrollo de aplicaciones haciendo uso de la tecnología Java, proporciona una arquitectura de aplicación fiable y flexible, es un producto libre y gratuito sin restricciones de uso ya que tiene una gran comunidad de usuarios y desarrolladores de todo el mundo.

El algoritmo está implementado en una arquitectura de tres capas: Presentación, lógica del negocio, y datos, en un nivel porque la forma física en que se encuentra distribuida reside en un solo ordenador. Tiene como propósito identificar las partes que intervienen en el desarrollo, y que darán el resultado esperado.

En la capa de datos residen los datos y es la encargada de acceder a los mismos, aquí se encuentra alojada la data de Twitter en una base de datos relacional MYSQL que es la que se trabajó en esta investigación, y mediante la conexión a la BD se permitirá el tratamiento y manipulación de información, mediante consultas que se deben procesar para obtener la información necesaria.

En la capa lógica se encuentra las funciones que se ejecutan, recibe las peticiones del usuario, para procesar la información y enviar las respuestas tras el proceso, se busca tweets que compartan una misma url y se recupera de cada recurso todas las menciones, etiquetas, urls cortas, y el texto, para luego determinar el número de cambios que existen entre ellos.

Y en la capa de presentación el sistema interactúa con el usuario y viceversa, mostrándole la aplicación para que sea ejecutada, y presente los resultados, como: el número total de cambios en menciones, etiquetas, urls cortas, y en el texto además se presenta el promedio general de estas métricas.

## 5 Resultados

En la tabla de resultados se puede observar que la mayor cantidad de cambios realizados sobre un tweet tiende a ocurrir sobre el texto o el contenido dentro del mismo, es decir, que a pesar de que las personas lleven a cabo modificaciones sobre el texto de un tweet tienden a conservar los rasgos clasificatorios más importantes como son las menciones, las etiquetas y las url que están dentro de dicho tweet.

**Tabla 1.** Promedio de los resultados de todas las métricas.

Muestra	Núm. tweets	Promedios en Menciones	Promedios de etiquetas	Promedio de urls cortas	Promedios del Texto
1	2500	2,202480992	3,370948379	5,835934374	4,93557423
2	5000	2,821643287	5,026052104	7,651302605	7,09218437
3	7500	1,993465795	2,811708228	4,632217629	4,32044273
4	10000	1,936693669	2,679467947	4,389238924	4,42154215
5	12500	1,927354188	2,654452356	4,243939515	4,55388431
6	15000	1,914927662	2,651243416	4,128208547	4,61184079
7	17500	1,91982399	2,626035773	4,024287102	4,61420653
8	20000	1,916645832	2,603530177	3,936346817	4,58257913
<b>TOTAL</b>	<b>22000</b>	<b>1,902177372</b>	<b>2,588799491</b>	<b>3,860493659</b>	<b>4,53879722</b>

## 6 Conclusiones

Luego de haber analizado una cantidad aproximada de 22000 tweets, se puede observar que los cambios realizados sobre cada tweet son más notables con respecto al contenido modificado dentro de los mismos.

En los resultados se puede observar que los promedios de cambios con respecto a las distintas menciones, etiquetas, y urls cortas tienden a disminuir conforme se aumenta el tamaño de la muestra analizada, puesto que los retweets tienden a ser replicados igual que el original y mientras mayor sea la población analizada menor es la probabilidad de encontrar cambios.



Los promedios de los cambios obtenidos de los textos, determinan que a medida que aumenta la muestra, también aumenta el número de cambios, lo que significa que los tweets originales son modificados en su mayoría con respecto al texto, y mientras mayor sea la población analizada mayor es la probabilidad de encontrar cambios.

## 7 Trabajos Futuros

Se puede emplear el estudio de cambios en los tweets para determinar las tendencias que seguirán un conjunto específico de datos o usuarios, este trabajo emplea análisis cuantitativos para determinar los cambios llevados a cabo sobre un tweet; en el caso de que se quiera ampliar el campo de investigación de la interacción en redes sociales, se puede buscar los cambios realizados en una muestra de tweets, pero ya enmarcando el análisis sobre un campo en específico.

Por ejemplo, se pueden llevar a cabo valoraciones y testeos de cambio en tweets que hablen sobre biología u otros temas, esto con el fin de establecer el auge y la influencia de un tema en específico dentro de una organización, empresa o comunidad virtual, al hacer esto se tendrá una idea aproximada de que datos son los más relevantes al momento de realizar una investigación o proyecto.

También mediante el análisis de cambios se pueden establecer las bases para diseñar, desarrollar e implementar recomendadores o filtros de información para búsqueda de información relevante dentro de una red social, enmarcándolo en un ámbito específico.

## Bibliografía

- Twitter. (2013). Obtenido de <http://www.twitter.com/>
- Carter, S., Tsagkias, M., & Weerkamp, W. (2011). *Twitter hashtags: Joint Translation and Clustering*. Obtenido de [http://www.websci11.org/fileadmin/websci/Posters/125\\_paper.pdf](http://www.websci11.org/fileadmin/websci/Posters/125_paper.pdf)
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (s.f.). *Measuring User Influence in Twitter: The Million Follower Fallacy*. Obtenido de Max Planck Institute for Software Systems (MPI-SWS), Germany; Royal Veterinary College, University of London, United Kingdom; CS Dept., Federal University of Minas Gerais (UFMG), Brazil: <http://www.google.com.ec/url?sa=t&rct=j&q=measuring%20user%20influence%20in%20twitter%3A%20the%20million%20follower%20fallacy&source=web&cd=1&ved=0CC8QFjAA&url=http%3A%2F%2Fwww.aaai.org%2Focs%2Findex.php%2FICWSM%2FICWSM10%2Fpaper%2Fdownload%2F1538%2F1826&>
- Collado C, E. (2011). *Máster universitario de Sociedad de la Información y el Conocimiento*. Obtenido de <http://eduangi.com/wp-content/uploads/2011/07/El-microblogging-en-el-periodismo-esp%C3%B1ol.pdf>
- De la Rosa Troyano, F. *Sistemas de Inteligencia Web: Análisis de Redes Sociales*. Obtenido de <http://www.lsi.us.es/~ffrosat/publicaciones/Tesis.2011.v75.vf.pdf>

- iLifebelt. (2013). *Estudio Redes Sociales Centroamérica 2013*. Obtenido de II Estudio de las Redes Sociales en América Central: <http://octavioislas.files.wordpress.com/2013/08/redes-sociales-centroamerica-2013.pdf>
- Iribarren, J. (17 de Julio de 2009). *Impact of Human Activity Patterns on the Dynamics of Information Diffusion*. Obtenido de IBM Corporation, ibm.com e-Relationship Marketing Europe, 28002 Madrid, Spain: <http://prl.aps.org/abstract/PRL/v103/i3/e038702>
- Jiménez , M. (2013). *Visualización de redes de colaboración a partir de datos abiertos en twitter aplicando técnica*. Obtenido de <http://dspace.utpl.edu.ec/bitstream/123456789/6595/1/Jimenez%20Mercy%20Elizabeth.pdf>
- Orihuela, J. (s.f.). *Profesor en la Facultad de Comunicación de la Universidad de Navarra y autor del blog eCuaderno*. Obtenido de <http://www.virtualoglos.net/portal/virtual-community/virtual-teens/pliteraria/147-twitter-como-medio-de-comunicacion.html>
- Orta, C., Pardo, F., & Salas, J. (s.f.). *Análisis de Redes de Influencia en Twitter*. Obtenido de <http://www.google.com.ec/url?sa=t&rct=j&q=an%C3%A1lisis%20de%20redes%20de%20influencia%20en%20twitter&source=web&cd=1&ved=0CDsQFjAA&url=http%3A%2F%2Fwww.redesymediossociales.es%2Fdescarga.php%3Ffile%3DAnálisis%2520de%2520influencia%2520en%2520Twitter.pdf&>
- Ponce, I. (17 de Abril de 2012). (17/4/2012): “ *Monográfico: Redes Sociales*”. *Observatorio Tecnológico NIPO:820-10-289-9*. Obtenido de <http://recursostic.educacion.es/observatorio/web/es/internet/web-20/1043-redes-sociales>
- Romero, D., Meeder, B., & Kleinberg, J. (Junio de 2011). *"Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter"*. Obtenido de (ed.) International World Wide Web Conference Committee (IW3C2): <http://www.cs.cornell.edu/home/kleinber/www11-hashtags.pdf>
- Ugarte, D. d. (s.f.). *Breve Historia del análisis de redes sociales*. Obtenido de [http://lasindias.com/gomi/historia\\_del\\_analisis\\_de\\_redes\\_sociales.pdf](http://lasindias.com/gomi/historia_del_analisis_de_redes_sociales.pdf)
- Weng, L., Menczer , F., & Ahn, Y. (28 de August de 2013). *Scientific reports. Virality Prediction and Community Structure in Social Networks*. Obtenido de Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA.: <http://www.nature.com/srep/2013/130828/srep02522/pdf/srep02522.pdf>
- Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., y otros. (20 de Junio de 2013). *The Role of Information Diffusion in the Evolution of Social Networks*. Obtenido de <http://arxiv.org/pdf/1302.6276v2.pdf>