

**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

***La Universidad Católica de Loja***

**ÁREA TÉCNICA**

**INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**

**BASE DE DATOS**

**Por:**

Joan Manuel Briceño Quilambaqui

Alex Fernando Aguirre Rojas

**Director:**

Audrey Elizabeth Romero Peláez

**Tema:**

Análisis del Proyecto Final

Loja – Ecuador

Enero – 17 – 2019

# Solución

*¿Cómo se relaciona la base de datos tweets con ods.csv?*

La relación fundamental entre el archivo csv y la base de datos, es principalmente que después de una análisis y limpieza de este este puede ser trabajado y subido en la base de datos, mediante Oracle SQL Developer se crean tablas, las cuales se van a ir agregando los datos que están en el dataset.

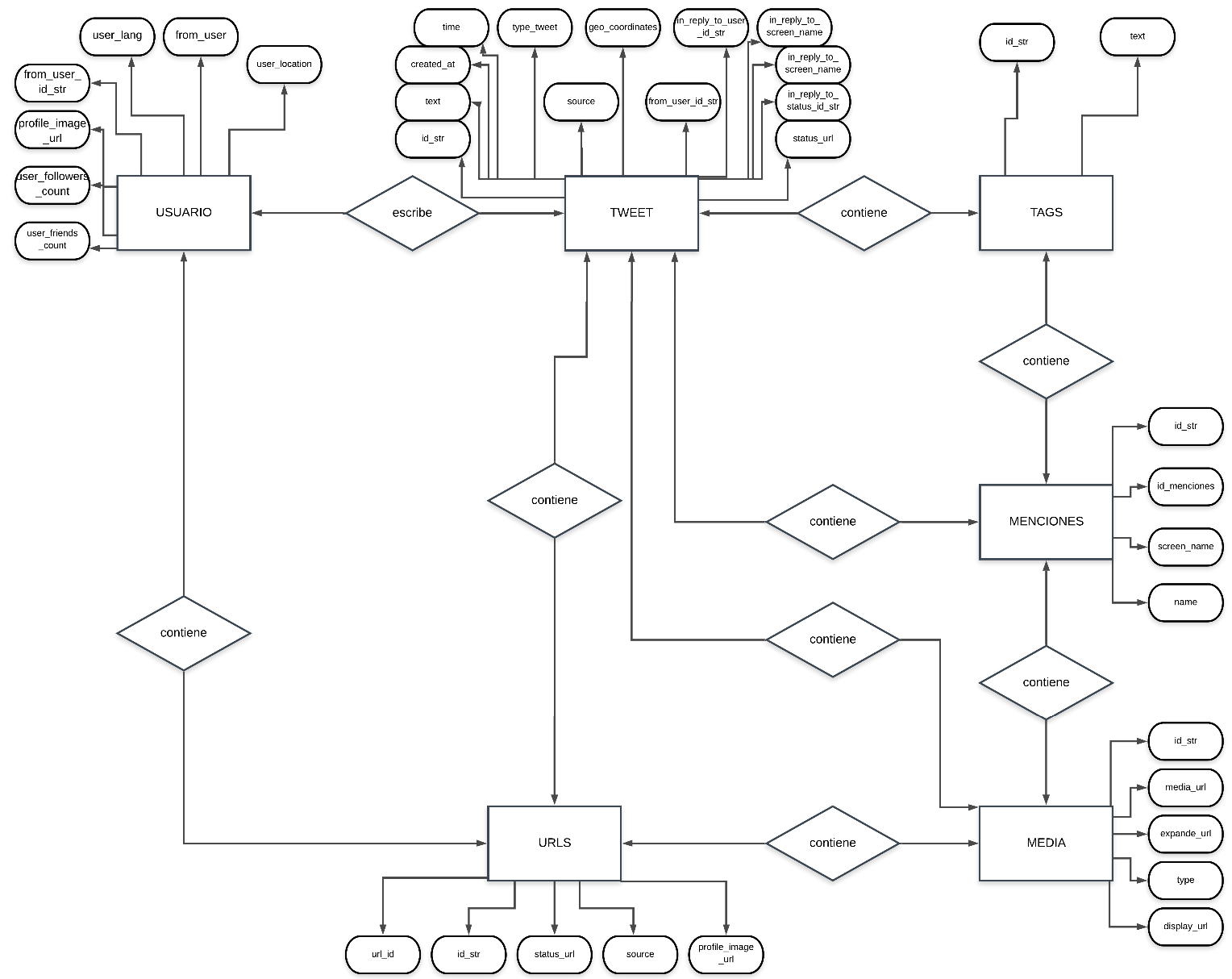
Los datos obtenidos del archivo en cuestión se representan en el siguiente diccionario:

Tabla 1 Diccionario de datos

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PK/FK** | **Columna** | **Tipo** | **Descripción** | **Tamaño** | **Ejemplo** |
| PK | id\_str | VARCHAR2 | Representa al identificador único de un Tweet | 30 | 1038336680929964032 |
|  | from\_user | VARCHAR2 | El nombre del usuario, tal como lo ha definido | 30 | MarielaAyerve |
|  | Text | VARCHAR2 | El texto, en UTF-8, de la publicación. | 280 | RT @Salud\_CZ7: #Loja | Cristian Abendaño, psicólogo clinico, aborda sobre la prevención del suicido. Enfatiza como factores de protección la.comunicación, resiliencia, habilidades sociales, tiempo de calidad @lahoraecuador https://t.co/ CCAw6qB6SD |
|  | created\_at | VARCHAR2 | Fecha UTC cuando se creó este Tweet | 30 | Sat Sep 08 08:02:10 +0000 2018 |
|  | Time | DateTime | Fecha |  | 08/09/2018 08:02:10 |
|  | geo\_coordinates | VARCHAR2 | Puede ser nulo. Representa la ubicación geográfica de este Tweet según lo informado por el usuario o la aplicación del cliente. | 50 | loc: 18.501,-69.917 |
|  | user\_lang | VARCHAR2 | Idioma con el que el usuario configuró su cuenta. | 6 | es |
|  | in\_reply\_to\_user\_id\_str | VARCHAR2 | Puede ser nulo. Si el Tweet es una respuesta, este campo contendrá la el ID del autor del Tweet original. Esto no siempre será necesariamente el usuario mencionado directamente en el Tweet. | 30 | 6253282 |
|  | in\_reply\_to\_screen\_name | VARCHAR2 | Puede ser nulo. Si el Tweet es una respuesta, este campo contendrá el nombre de pantalla del autor original del Tweet | 30 | twitterapi |
|  | from\_user\_id\_str | VARCHAR2 | La cadena representa al identificador único para este usuario | 30 | 915406661157818368 |
|  | in\_reply\_to\_status\_id\_str | VARCHAR2 | Puede ser nulo. Si el Tweet es una respuesta, este campo contendrá la representación de cadena del ID del Tweet original. | 30 | 1051222721923756032 |
|  | source | VARCHAR2 | Aplicación utilizada para publicar el Tweet. (Formato HTML) | 30 | <a href="http:// twitter.com/download/ android" rel="nofollow">Twitter for Android</a> |
|  | profile\_image\_url | VARCHAR2 | URL que apunta a la imagen de perfil del usuario | 100 | http://pbs.twimg.com/ profile\_images/ 915452554905038848/1 rVz0aNX\_normal.jpg |
|  | user\_followers\_count | NUMERIC | El número de seguidores que esta cuenta tiene actualmente. Bajo ciertas condiciones de coacción, este campo indicará temporalmente "0" | 1000 | 41 |
|  | user\_friends\_count | NUMERIC | El número de usuarios que esta cuenta está siguiendo (también conocido como sus "seguidores"). Bajo ciertas condiciones de coacción, este campo indicará temporalmente "0" | 1000 | 83 |
|  | user\_location | VARCHAR2 | Puede ser nulo. La ubicación definida por el usuario en el perfil de su cuenta. | 30 | San Francisco, CA |
|  | status\_url | VARCHAR2 | URL que apunta al tweet original | 100 | http://twitter.com/ MarielaAyerve/statuses/ 1038336680929964032 |

**Modelado Conceptual**

El modelo conceptual es un primer paso para la construcción del modelo lógico, se debe tomar en cuenta que el id\_str es la clave primaria de la base de datos, las demás tablas de no poseen cables primaria puesto a que en una base de datos sol existe una sola, las claves foráneas si son utilizadas por las tablas urls, media, hashtags, usuario y menciones.



**Descripción de entidades.**

**USUARIO:** El usuario es la persona u organización responsable de redactar y publicar un tweet. Algunos de los atributos más relevantes del usuario son el nombre que se muestra en pantalla, la foto de perfil, los seguidores, amigos, ubicación y su ID único de usuario que se relaciona con la tabla del tweet para indicar que usuario redacta el tweet.

**TWEET:** Esta entidad es la principal en la base de datos, todas están relacionadas a ella y esta contiene todo lo que forma parte de un tweet o un retweet, el tweet es la publicación realizada por un usuario que puede contener letras, números, signos y enlaces, pero con un límite de 280 caracteres. Además, cada tweet contiene un ID único, el texto por el que está conformado, hora y fecha, y en ocasiones la localización de donde fue publicado el tweet.

**TAGS:** Los tags o etiquetas son escritos con el sigo # antepuesto que se utiliza para escribir palabras clave o temas específicos en Twitter.

Estos tags los podemos encontrar en la columna de “entities\_str” que poseen un formato JSON que forman parte de los Datos No Estructurados por lo que no se los puede visualizar con facilidad, pero se pueden visualizar por medio de la aplicación web JSON VIEWER

La entidad TAGS cuenta con el atributo id\_str para que se relacione con la entidad principal RE/TWEET ya que los tags son dependientes del tweet.

**URLS:** El URL del tweet es la dirección especifica que se asigna a cada uno de los recursos disponibles en la red con la finalidad de que puedan ser localizados o identificados. La URL del tweet es distinta a la de la media debido a que esta entidad hace referencia directamente al tweet especifico. Se relaciona directamente con el tweet por lo que la clave foránea como en los demás casos es el id\_str, dependiendo directamente del tweet.

**MENCIONES:** Las menciones sirven para interactuar con otros usuarios y contiene el nombre del usuario que se quiere mencionar con la arroba (@) antepuesto, el ID de la mención y debido a que esta entidad también procede del JSON se utiliza la misma clave foránea id\_str para relacionar con la entidad tweet.

**MEDIA:** Esta entidad hace referencia a los elementos multimedia pueden ser utilizados un tweet y contiene un atributo tipo que especifica si es una imagen, video, gif, etc.; además del URL de la media. Estos son extraídos del JSON.

**Modelado Lógico**

En el apartado de modelo lógico se da a cada atributo un dominio, se colocan las claves primarías y foráneas en las tablas; se reevalúan las relaciones y si es necesario se eliminan las conexiones.

En el siguiente grafico se puede ver las relaciones que tiene las tablas entre sí, cada entidad tiene sus propios atributos los cuales se relacionan principalmente con tweet.

