Trabajo Final. Administración de base de datos en MySQL

# Tabla de Contenido

Contenido

[Tabla de Contenido 1](#_Toc142230381)

[Proyecto Base de Datos Repartidores. 1](#_Toc142230382)

[Innovación y modelo de negocios 2](#_Toc142230383)

[Las características estructurales de la base de datos 3](#_Toc142230384)

[VIEWS 3](#_Toc142230385)

[FUNCIONES 4](#_Toc142230386)

[PROCEDIMIENTOS 4](#_Toc142230387)

[TRIGGERS 5](#_Toc142230388)

[CONTROL DE TRANSACCIONES 5](#_Toc142230389)

[CONTROLES EN LA ADMINISTRACION DE LA BASE DE DATOS 6](#_Toc142230390)

# Proyecto Base de Datos Repartidores.

Las técnicas de análisis de la conflictividad social han tenido una importante revolución a partir del desarrollo de herramientas computacionales aplicadas en la investigación en ciencias sociales. Entre ellas, el rascado (scraping por su denominación anglosajona) de periódicos digitales permite obtener abundante información sobre conflictos sociales, como huelgas y movilizaciones de trabajadores, y sobre sujetos en particular, como sindicatos o colectivos de trabajadores, entre otros. En general, uno de los principales objetivos que se persigue con esta técnica es la construcción de una base de datos que contenga toda la información posible de una temática de interés y que estructure la información de la siguiente manera: link de la nota, fecha, sección, título de la noticia y texto de la noticia.

Posteriormente se emplean técnicas de minería de texto que permiten extraer mayor información del cuerpo de la noticia. Esta información es seleccionada según el interés del proyecto de investigación. En nuestro caso trabajamos sobre la conflictividad entre repartidores y empresas de plataformas como Rappi y Pedidos YA y las políticas públicas en torno a la regulación laboral o comercial de la actividad de los repartidores. El registro de los hechos de conflictividad nos permite medir a lo largo del tiempo y del espacio como se desarrollan los enfrentamientos entre trabajadores y empresas, que repertorio de protesta tiende a predominar, que demandas son más recurrentes y que organizaciones tienden a agrupar a los trabajadores. A su vez, el proceso de regulación busca proteger a los trabajadores de un proceso de trabajo que los vuelve vulnerables frente a los accidentes en la vía pública y los hechos delictivos. Los periódicos nacionales han notado el interés público sobre esta problemática y han dado cobertura tanto a accidentes de tránsito fatales como a los robos violentos. Por este motivo, podemos medir a lo largo del tiempo la cantidad de notas que mencionan hechos luctuosos para la vida del repartidor. Finalmente, nos interesa medir las políticas estatales favorables y desfavorables a la regulación a la regulación laboral de la actividad. De esta forma, podremos conocer en qué periodo de tiempo y que instituciones o partidos políticos fueron más activos en torno al establecimiento de un contrato comercial o laboral.

La creación de una base de datos de estas características nos permitiría reunir una abundante información sobre el sector, realizar mediciones y buscar establecer correlaciones entre las distintas variables. A través de esta base de datos podremos conocer las principales reivindicaciones de los trabajadores, los repertorios de protesta más empleados, los resultados de los enfrentamientos entre los repartidores y las empresas, la inseguridad en el proceso de trabajo y las políticas en torno a la regulación de la actividad.

Por las características de la base de datos nuestra tabla primaria es la «tabla fuente» que reúne toda la información de la noticia en cinco columnas: *link*, *fecha*, *sección*, *título* y *texto*. Esta tabla es la principal porque en la columna texto se agrupa toda la información de la noticia y permite, en caso de ser necesario, un análisis de mayor profundidad.

De la tabla madre –llamada fuente- se desprende una tabla hija que agrupa las fechas de las notas y los lugares en donde se produjeron los hechos registrados: país, provincia, ciudad. A su vez, como puede ocurrir que una noticia refiera a muchos lugares, por ejemplo, un paro que es acatado por repartidores de distintas ciudades, de esta tabla se desprende una tabla hija que permite cargar mayor cantidad de lugares.

Hay dos tablas (hijas) que refieren a hechos luctuosos y que se desprenden de la tabla madre Fuente: la tabla «accidente de tránsito» y la tabla «inseguridad». La primera tabla registra el *tipo de accidente*, *el vehículo* en que se transportaba el repartido (bici o moto) y la *causa (*colisión o lluvia, por ejemplo*)*. De estas dos tablas se crea una tabla hija llamada «salud del repartidor» que registra la cantidad de repartidores que sufrieron accidentes de tránsito o robos violentos y su estado de salud. Se registra la información en las columnas «estado», «obra social», «ART».

De la tabla fuente se desprende una tabla que agrupa la conflictividad en el sector. La Misma agrupa la cantidad de conflictos realizados y permite analizar su evolución a lo largo del tiempo. De esta tabla, se desprenden cuatro tablas hijas nombradas «Demandas», «Acciones», «Organización de Trabajadores» y «Empresa» que agrupan información específica del conflicto. La tabla Demanda agrupa en sus columnas el Tipo de Demanda (Salarial, Seguridad Social, Inseguridad, Reconocimiento Laboral) y la Demanda concreta. La Tabla Acción reagrupa información acerca de los actos realizados por los trabajadores o por la empresa en los distintos hechos conflictivos, está integrado por las columnas Tipo de Acción (movilización, paro, desconexión, despido, Lock Out), Acción (con el nombre concreto y una descripción) y Participantes (la cantidad de estos). La Tabla Organización brinda información de la asociación de trabajadores que llevaron a cabo las acciones, el Nombre concreto, el Tipo de Organización (sindicato tradicional, agrupación, red de trabajadores) y la Orientación (clasista, peronista, apolitica). Por último, la Tabla Empresa registra información de las empresas que participaron de los conflictos en las distintas etapas.

Por último, de la tabla fuente se desprende la tabla llamada «políticas públicas» que registran las acciones estatales llevadas a cabo por distintos poderes estatales, instituciones y partidos políticos que intervinieron en el escenario de conflictividad y dieron forma a la regulación estatal. En esta tabla se administra información de las Instituciones intervinientes, las Acciones –políticas publicas llevadas a cabo-, y los Objetivos que perseguía el Estado con dichas intervenciones.

# Innovación y modelo de negocios

El «modelo de negocios» de este registro se vincula a las necesidades estatales de obtener la mayor cantidad de información posible sobre un sector de trabajadores con el objetivo de organizar las políticas públicas de la forma más eficaz posible. ¿de qué forma regular la actividad de los repartidores? ¿Cuáles son los problemas que enfrenta el sector? Son preguntas que el Estado ha resuelto a lo largo del tiempo, pero que ahora en las nuevas modalidades de empleos por plataforma pueden requerir o no, nuevas soluciones. Conocer el conflicto es una forma de conocer las demandas de estos trabajadores y de las empresas con el fin de establecer soluciones posibles.

En este caso ofrecemos un modelo de base de datos experimental que nos permita comenzar a reflexionar como las nuevas herramientas computacionales pueden ahorran presupuesto estatal registrando automáticamente la información de periódicos digitales. En este trabajo se propone un esquema de base de datos que puede ser perfeccionado en un futuro, por ejemplo, automatizando la carga de datos de la tabla madre a las tablas hijas a partir del análisis de lenguaje en el cuerpo de la nota y el título.

# Las características estructurales de la base de datos

La base de datos y sus distintos objetos son el resultado del *rascado* de noticias de periódicos nacionales, en este caso Infobae. Por este motivo, la mayoría de los datos con los que contamos son textuales y fueron establecidos en las distintas tablas como TEXT, CHAR o MEDIUM TEXT, VARCHAR, según la extensión de la cadena de texto o la posibilidad de encontrar otro tipo de caracteres como números en los cuerpos de la nota. Empleamos ENUM en los casos que podíamos cargar opciones pre establecidas como el “tipo de accion” cuando nos interesaba mantener una tipificación particular. En pocas ocasiones trabajamos con datos numéricos, lo hicimos fundamentalmente para el registro de las claves primarias y foráneas, en todos os casos como INT y cuando registramos participación de huelguistas como SMALLINT.

La estructura de la base de datos tiende a tener una estructura vertical dado que toda la información se desprende de la tabla fuente que concentra la información de una noticia periodística. De esta fuente surgen tres segmentos de análisis: inseguridad/accidentes de tránsito, conflictos y políticas públicas. Los objetivos inmediatos de la administración de la base de datos son mostrar información relevante de estos tres segmentos y realizar análisis sobre los cuerpos textuales de las noticias para establecer mediciones significativas, a partir –por ejemplo-, de palabras claves.

Proponemos la siguiente estructura de informes:

# VIEWS

Una View –vista- es una consulta almacenada que se trata como una tabla virtual. Permite a los usuarios acceder y manipular datos sin necesidad de acceder directamente a las tablas subyacentes, simplificando consultas complejas o proporcionando una vista personalizada de los datos. En este caso realizamos cinco vistas:

* La primera, se denomina ‘violencias’ y permite al usuario consultar el estado de los repartidores que sufrieron accidentes o robos violentos;
* La segunda, nombrada ‘participantes’ muestra las acciones colectivas que agruparon a más de 100 repartidores y las agrupaciones que impulsaron dichas acciones;
* La tercera, se denomina ‘huelgas’. Busca en la base de datos aquellos titulares que refieran a los hechos conflictivos huelgas o protesta y permite al usuario conocer el titulo, el id. De la nota y la fecha de la misma;
* La cuarta, es ‘conflictividad’ reagrupa información relevante de distintas tablas que agrupan noticias sobre conflictos laborales. El usuario podrá conocer el tipo de demanda, la demanda concreta, el tipo de acción colectiva empleada, la acción concreta, el tipo de organización y su nombre.
* La quinta, se nombra ‘acción judicial’ muestra todas las menciones en donde figura que el poder judicial realiza una acción que promueve una regulación laboral de la actividad.

# FUNCIONES

Una función es un bloque de código reutilizable que realiza operaciones específicas en los datos y devuelve un resultado. Puede aceptar parámetros de entrada y generar un valor de salida, facilita la ejecución de tareas repetitivas en consultas y procedimientos almacenados. En este caso, vamos a ofrecer al usuario tres funciones distintas que permiten “estimar” huelguistas totales y medir la relevancia de ciertas acciones en los corpus textuales de las noticias.

En el primer ejercicio, se crea la función “f\_huelguistas” que mide dos variables que son cantidad de acciones (movilizaciones o paros, por ejemplo) y las multiplica por una cantidad de huelguistas determinada con el objetivo de conocer el total de huelguistas[[1]](#footnote-1).

La segunda función “fn\_palabra\_t” le permite al usuario realizar una medición por la que se conoce la relevancia de determinada palabra clave en relación a la cantidad de noticias. Esta medición permite al usuario establecer la relevancia de una organización, una demanda o una acción en el conjunto de las notas periodísticas. La medición se establece a partir de contar la cantidad de veces que se menciona dicha palabra en la columna "titulo" y posteriormente se divide dicha cantidad por el total de noticias (filas).

La tercera función “fn\_palabra\_c” es equivalente a la anterior solo que en vez de contar la frecuencia de palabras en la columna título lo realiza en la columna texto, donde se encuentra el contenido completo de la noticia periodística.

# PROCEDIMIENTOS

Un procedimiento es un conjunto de instrucciones SQL predefinidas que pueden aceptar parámetros de entrada, ejecutar operaciones en la base de datos y, opcionalmente, generar resultados o realizar acciones específicas. Los procedimientos almacenados permiten encapsular lógica de negocio y tareas complejas en una sola unidad, lo que facilita su reutilización y mantenimiento. En este caso ofrecemos al usuario tres procedimientos.

* El primero permite filtrar la información de la tabla ‘politicas publicas’. El procedimiento selecciona las filas en las que una palabra clave cargada por el usuario se encuentra en la columna de objetivo.
* El segundo permite contabilizar la cantidad de filas que refieren a tipos de acción de la conflictividad social.
* El tercero, genera una tabla que permite ordenar la tabla ‘acción’ a partir de una palabra clave

# TRIGGERS

Un Trigger en MySQL es un objeto de base de datos que se activa automáticamente en respuesta a ciertos eventos, como la inserción, actualización o eliminación de datos en una tabla específica. Cuando se produce el evento especificado, el trigger ejecuta un conjunto predefinido de instrucciones SQL, lo que permite automatizar acciones o aplicar lógica personalizada en la base de datos en tiempo real. Los triggers son útiles para implementar reglas de negocio, mantener integridad de datos y realizar auditorías. En este caso ofrecemos al usuario tres triggers.

* El primero copia la información cargada por un usuario en la tabla ‘politicas\_publicas’ en una tabla equivalente.
* El segundo, copia cierta información relevante de la tabla ‘politicas\_publicas’ en una tabla de auditoria que agrega el día de la modificación de datos y el usuario que realizo dicha modificación.
* La tercera, copia cierta información relevante de la tabla ‘politicas\_publicas’ en una tabla de auditoria que informa el usuario que realizo la modificación, el tipo de acción que realizo (eliminación, actualización o carga de datos) y el día en que hizo la acción.

# CONTROL DE TRANSACCIONES

El Lenguaje de Control de Transacciones (TCL, por sus siglas en inglés) se utiliza en SQL para administrar y controlar las transacciones en una base de datos. TCL incluye comandos como COMMIT (confirmar), ROLLBACK (deshacer) y SAVEPOINT (punto de guardado), que permiten a los desarrolladores y administradores de bases de datos gestionar la integridad y consistencia de las transacciones, asegurando que los cambios en los datos se realicen de manera controlada y confiable.

En el documento lenguaje TCL programamos todas estas funciones en la base de datos.

# CONTROLES EN LA ADMINISTRACION DE LA BASE DE DATOS

El Lenguaje de Control de Datos (DCL, por sus siglas en inglés) se utiliza en SQL para administrar permisos y controlar el acceso a los datos en una base de datos. DCL incluye comandos como GRANT (conceder permisos) y REVOKE (revocar permisos), que permiten a los administradores de la base de datos otorgar o retirar privilegios a usuarios y roles, asegurando la seguridad y la gestión adecuada de los datos.

En el trabajo generamos y revocamos distintos permisos a usuarios que gestionan la base de datos.

1. El ejercicio tiene un límite claro dado que presupone que siempre la cantidad de huelguistas es equivalente en todas las movilizaciones. De todas formas, el objetivo es comenzar a probar el funcionamiento de las funciones y en este sentido arroja un resultado satisfactorio. [↑](#footnote-ref-1)