

Normativa

1. La práctica se realizará en equipos con un mínimo de 3 alumnos y un máximo 5 alumnos. No obstante, el tamaño final de los equipos será fijado por el profesor de prácticas en función de las necesidades académicas de cada grupo de matrícula.
2. Todos los integrantes del equipo de prácticas tienen que estar **matriculados en el mismo grupo de clase**.
3. La práctica consiste en 3 apartados distintos: 1) modelado del sistema. 2) creación y explotación de la base de datos y 3) programación contra la base de datos.
4. La evaluación de la práctica se realizará en el aula por parte del profesorado de la asignatura durante la última semana de curso (semana del 18 de diciembre). Es obligatoria la presencia en el aula y responder a las preguntas planteadas por el profesor el día de la evaluación. Los alumnos que no estén presentes recibirán una calificación de 0.0. La práctica se evaluará una única vez.
5. Los profesores se reservan el derecho a realizar una evaluación individualizada de la práctica a los estudiantes de la asignatura con el fin de comprobar la participación de cada estudiante en el grupo de prácticas.
6. Los resultados de la realización de la práctica deberán ser entregados a través de Moodle antes de su evaluación en un formato electrónico.
7. De acuerdo con la Normativa de Evaluación de la Universidad Politécnica de Madrid, la comprobación de fraude académico durante la realización de la práctica supondrá la obtención de una calificación de cero en la misma y podrá implicar consecuencias más severas en función de la gravedad de los hechos.

Enunciado

Un conjunto de editoriales que trabajan coordinadamente desean automatizar algunos de los procesos del tratamiento de las revistas que publican, junto con sus autores y datos más relevantes. Estas **revistas** están clasificadas según su **factor de impacto**, que es un número que describe, de alguna manera, la importancia que una publicación tiene para la comunidad científica (generalmente a través del número de citas que se hace de cada uno de sus **artículos** en otros artículos). Este factor de impacto puede variar a lo largo del tiempo pero en este caso se va a trabajar con un único valor asociado a cada publicación. Estas publicaciones se ordenan de acuerdo con este factor de impacto conformando una lista que se divide en cuartiles (cuatro divisiones en la lista), siendo las revistas del primer cuartil las que mayor valor de factor de impacto presentan.

Inicialmente se van a procesar solamente un conjunto de artículos y autores europeos de unos determinados años y para ello nos solicitan que diseñemos una base de datos que contemple las siguientes restricciones:

Cada una de las **editoriales** se caracteriza por un **nombre** y una **nacionalidad** principal de la misma. Cada una de ellas **publica, de manera periódica** (que hay que conocer y que puede ser distinta por revista), un conjunto de revistas (**journal**) que a su vez se caracterizan por un identificador único (**journal_id**), un nombre (**journal_name**), su **issn** (es un número único asociado a cada revista) y su factor de impacto asociado (JIF o Journal Impact Factor). Además, y aunque se sabe que el cuartil al que pertenece la revista se calcula en función de su factor de impacto, por motivos de eficiencia, el conjunto de editoriales desea guardar de forma explícita este dato (**JIF_Quartile**).

Por supuesto cuando un **autor** (o conjunto de autores), identificado por un **número entero** y su **nombre completo** con el que **firma los artículos** (la firma en un artículo generalmente incluye el primer apellido y la inicial del nombre separados por una coma, pero puede variar según las costumbres) desea publicar un artículo se lo **envía al editor** de la revista para que comience el proceso. Este proceso se inicia con el envío del artículo, por parte del **editor**, a un conjunto de otros autores diferentes (revisores a partir de ahora) que trabajen en el mismo área de investigación que el/los autores del artículo para que lo revisen, comprobando su corrección y validez científica. Un determinado profesional puede ser **autor de un conjunto de artículos** y ser **revisor de otros**, siempre dentro del área de investigación o temática en la que trabaja. Realmente, los editores no necesitan clasificar a los autores en una determinada temática para poder seleccionarlos como revisores de algún artículo ya que consideran que un autor trabaja en una determinada área si ha publicado algún artículo en esa área y por tanto, con clasificar los artículos por áreas de investigación es suficiente ya que quedan automáticamente clasificados sus autores. De hecho, **todos los artículos de una misma revista pertenecerán al misma área de investigación**, al estar éstas centradas en un tema. Al enviar un artículo para revisión, es absolutamente necesario conocer a qué revisores ha sido enviado, y para cada uno de los revisores, en qué **fecha se le ha enviado**, si dicha revisión ha sido **realizada**, en qué **fecha ha sido realizada** y si el **resultado de la revisión** ha sido positivo o no (un revisor puede recomendar aceptar o rechazar cada uno de los artículos que revisa).

Hay que tener en cuenta que, de cada autor interesa conocer también su **afiliación**, esto es, la entidad/empresa para la que trabaja. Esta **afiliación** se identifica por un número entero que puede ser muy grande (**affiliation_id**), el nombre de la institución (**affiliation_name**), la ciudad asociada a dicha afiliación (**city**) y el país al que pertenece (**country_name**). Un autor puede tener más de una institución asociada al mismo tiempo y por supuesto muchos pueden pertenecer a la misma entidad.

Los datos de los artículos que se desean almacenar en las base de datos, asociados, por supuesto, a sus autores (un artículo puede tener un número indeterminado de autores pero siempre tendrá al menos uno) son el **DOI** o Digital Object Identifier (que es una cadena alfanumérica que identifica de manera única a cada artículo a nivel mundial y que puede tener hasta 50 caracteres), el título del artículo (**title**), la fecha de publicación (**publication_date**) si se ha producido dicha publicación, la **url** en la que está disponible el artículo y el número de citas que tiene el artículo en otras publicaciones (**num_citations**). Cada uno de estos artículos debe estar **asociado con la revista** a la que se envía para su publicación, ya que el factor de impacto que afecta al artículo es el de la revista.

Todos los autores de un artículo deben asociarse al mismo, sin hacer distinción entre ellos. Sin embargo, se desea representar también el hecho de que algunos de estos autores forman **grupos de trabajo** en un determinado área de investigación (un área de investigación dispone de un **identificador numérico único** y un **nombre descriptivo** de la misma, por ejemplo: “*machine learning*”). Estos grupos, que no tienen relación alguna con los artículos publicados si no es por los autores, deben disponer de un identificador y un nombre, así como estar asociado a alguna de las entidades que figuran como afiliación de los autores. Un autor puede **pertenecer a más de un grupo** de trabajo y un mismo grupo de trabajo puede tener más de una afiliación ya que dependerá de la que asocie cada uno de los autores (pero solamente una por cada autor del grupo). Cada grupo ha sido creado para trabajar en una **única área de investigación** que debe figurar de manera obligatoria.

Además de lo anterior, el grupo de editoriales sabe que algunos de los autores, perteneciendo o no a alguno de los grupos anteriormente citados, colaboran entre ellos para poder llevar a término alguno de los artículos que un autor o autores está escribiendo. Sin embargo, esta colaboración no está sujeta a que los colaboradores figuren como autores de los artículos (puede ser una revisión previa o una corrección ortográfica del idioma, etc.). Es por ello que están también interesados en reflejar este hecho. La **colaboración**, que siempre tiene como objeto la elaboración de artículos, puede ser entre un número indeterminado de autores y

de manera independiente, para tantos artículos como deseen.

Por último, un artículo puede hacer **referencia** a otros y se necesita conocer cuáles hacen referencia a otros (estas referencias, a diferencia del número de citas, son solamente entre artículos que figuren en la base de datos).

Apartado 1

Completar las siguientes tareas:

1. Definir el modelado conceptual que cumpla con los requisitos escritos en el enunciado de la práctica. Será obligatorio utilizar el modelo Entidad-Relación con notación Chen, siendo motivo de calificación de NO APTO si se utiliza otro modelado.
2. Definir todos los elementos de “Semántica no contemplada” que se detecten.
3. Definir los dominios de todos los atributos que se incluyan en las entidades pertenecientes a los **artículos**, **revistas** y **autores**.
4. Realizar el paso a tablas del Modelo Entidad-Relación generado.

Apartado 2

Completar las siguientes tareas:

1. Define las instrucciones SQL para crear todas las tablas necesarias del paso a tablas que se realizó en el Apartado 1.4. Presta especial atención a las configuraciones de integridad referencial para garantizar que se cumplen todas las restricciones planteadas en el enunciado de la práctica.
2. El consorcio de editoriales desea, además, automatizar algunas consultas y nos pide que nos centremos en los procesos solamente de algunas tablas. Estas tablas son las que afectan a los artículos, revistas, autores y afiliaciones, con sus relaciones. Estas tablas han sido nombradas respectivamente: **article**, **journal**, **author**, **affiliation**, **author_affiliation** y **author_article**. Se pide¹:
 - a) Usando el fichero denominado **creardb2324.sql** disponible en Moodle, crear la base de datos **bd2324** y cargar todos los datos disponibles en dicho fichero en las tablas mencionadas y que van a ser objeto de procesos en puntos posteriores.
 - b) Estudiar, mediante ingeniería inversa, la estructura propuesta en este fichero, así como sus relaciones y atributos, y realizar el modelo relacional incluido en esta estructura.
 - c) Resolver en SQL la consulta: “Obtener los nombres, en orden alfabético, de aquellos autores que perteneciendo a la “Universidad Politécnica de Madrid” han publicado algún artículo en el año 2020 o en el año 2021”.

¹Se ha verificado que todas las consultas que se piden pueden ser realizadas con la base de datos proporcionada. Igualmente se garantiza que todas las consultas requeridas devuelven al menos un resultado con el conjunto de datos provisto. Aunque el tiempo de ejecución para todas consultas es “breve”, esto puede variar en función de la implementación realizada. Si una consulta tarda demasiado tiempo (por ejemplo, más de 5 minutos), es porque se está desarrollando una solución poco eficiente de la misma. Reflexionar sobre este hecho en caso de producirse para encontrar una solución adecuada.

- d) Resolver en SQL la consulta: “Obtener los nombres, en orden alfabético, de aquellos autores que perteneciendo a la “Universidad Politécnica de Madrid” han publicado algún artículo en el año 2020 y también en el año 2021”.
- e) Resolver en SQL la consulta: “Obtener el nombre de los autores y el nombre de la afiliación de aquellos autores que, perteneciendo a alguna entidad española, no han publicado ningún artículo ni en 2020 ni en 2021, ordenados por afiliación y dentro de cada entidad, por nombre de autor”.
- f) Resolver en SQL la consulta: “Obtener el nombre de la revista, su issn y el total de citas (*num_citations*) de todos los artículos publicados para cada una de ellas en aquellas revistas que están clasificadas dentro del primer cuartil de factor de impacto (*q1*)”.
- g) Resolver en SQL la consulta: “Obtener el nombre de la revista y el total de citas (*num_citations*) que hayan recibido sus artículos para aquella/s revista/s que, perteneciendo al primer cuartil del factor de impacto (*q1*), tengan el mayor número de citas de toda la base de datos”.
- h) Resolver en SQL la consulta: “Obtener aquellas entidades (*affiliation_name*) que tienen asociados al menos 10 autores que hayan publicado más de 5 artículos en el año más reciente que figuren en la base de datos (este año debe calcularse de forma dinámica con la consulta)”.
- i) Resolver en SQL la consulta: “Obtener el nombre de aquellas revistas que, habiendo publicado más de 300 artículos en el año de mayor antigüedad que figure en la base de datos tengan un factor de impacto (*jif*) superior a la media de los factores de impacto del global de las revistas de la base de datos. El año debe calcularse de forma dinámica con la consulta”.
- j) Resolver en SQL la consulta: “Obtener el nombre de aquellas revistas que hayan recibido en el global de sus artículos un mayor número de citas que la media de las citas recibidas por cada revista para el global de sus artículos de la base de datos”.
- k) Resolver en SQL la consulta: “Obtener el nombre de aquellos autores que hayan publicado artículos en todos los diferentes años que figuren en la base de datos”.
- l) Crear un procedimiento que, recibiendo un año como parámetro, devuelva, en dos parámetros de salida, el nombre de la revista y el número de autores para aquella revista que haya publicado en dicho año el artículo con un mayor número de autores. Si existen varias revistas con el máximo de autores, habrá que proporcionar como resultado una cadena que contenga, separados por puntos y comas, los nombres de dichas revistas (en un parámetro de salida solamente puede devolverse un valor). Se deberá usar obligatoriamente, al menos, un cursor en la resolución del procedimiento.
- m) Crear una función que, recibiendo como parámetro un identificador de una revista devuelva el número medio de artículos por año que dicha revista ha publicado. Escribir el código necesario para poder probarlo.
- n) Crear la tabla correspondiente a la asignación y realización de las revisiones de los artículos por parte de los autores (que actúan como revisores) que se ha obtenido en el análisis inicial. Una vez creada, crear un *trigger* que impida que se pueda insertar a un revisor de un artículo si dicho revisor figura ya como autor del mismo. Probar su funcionamiento con el código necesario.

Apartado 3

JDBC

Se deberá completar el desarrollo de tres métodos que se enunciarán a continuación, haciendo uso de MySQL Connector/J 8.0. Los métodos que se deben implementar tienen que comunicarse con la base de datos, que ya debería estar creada tras completar el Apartado 2:

1. Añadir a los integrantes del grupo como autores en la base de datos. Para ello complete previamente el método: `nuevoAutor(String authorName)`. Al no disponer de ninguna publicación, la importancia de los nuevos autores debe ser 0.
2. Mostrar por pantalla la lista de artículos (título y fecha de publicación) del autor “*Ortega F.*” en el año 2021. Para ello complete previamente el método: `listaArticulosPorAutor(String authorName, int year)`.
3. Mostrar por pantalla una lista con las instituciones y el número de autores que tiene cada una ordenadas de mayor a menor número de autores. Para ello complete previamente el método: `listaAfiliaciones()`.

Para llevar a cabo la implementación de los métodos anteriores, se deberá completar las siguientes tareas:

1. Descargar la plantilla de código fuente para completar este sub-apartado y cargarlo en un proyecto Maven de IntelliJ. El código fuente se encuentra en GitHub: <https://github.com/bbddetsisi/practica-bd2324-jdbc>
2. Inicialice el servidor de MySQL.
3. Complete las funciones proporcionadas en la clase `Main`. Para ello analice el código del proyecto y complételo siguiendo las instrucciones de los comentarios `@TODO`.

Hibernate

Para completar este sub-apartado se deberá completar las siguientes tareas:

1. Descargar la plantilla de código fuente para completar este sub-apartado y cargarlo en un proyecto Maven de IntelliJ. El código fuente se encuentra en GitHub: <https://github.com/bbddetsisi/practica-bd2324-hibernate>
2. Analice el fichero `hibernate.cfg.xml` para determinar el nombre del *schema* en el que hibernate creará la base de datos..
3. Realice las anotaciones en las clases `Affiliation`, `Article`, `Author` y `Journal` para que respeten las restricciones del Modelo Entidad-Relación en el que derivaría el modelo relacional obtenido en el apartado 2.2.2. Únicamente será necesario incluir el código de las anotaciones, el resto de código de las clases no necesita ser modificado.
4. Complete las instrucciones `@TODO` del método `main` de la clase `Main`. Tal y como se indica deberá:
 - a) Crear una afiliación con nombre “*Universidad Politécnica de Madrid*” de la ciudad de Madrid (España) y guardar dicha afiliación en la base de datos.

- b) Leer el fichero CSV **authors.csv** que se encuentra en el directorio **resources** y recorrerlo usando **CSVParser** para crear los autores que en el se encuentran. Dichos autores deberán ser asignados a la afiliación creada en el apartado anterior. Se deben guardar estos autores y sus afiliaciones en la base de datos.