



Análisis, Diseño y Procesamiento de Datos Aplicados a las Ciencias y a las Tecnologías

PRÁCTICAS (APARTADO III)

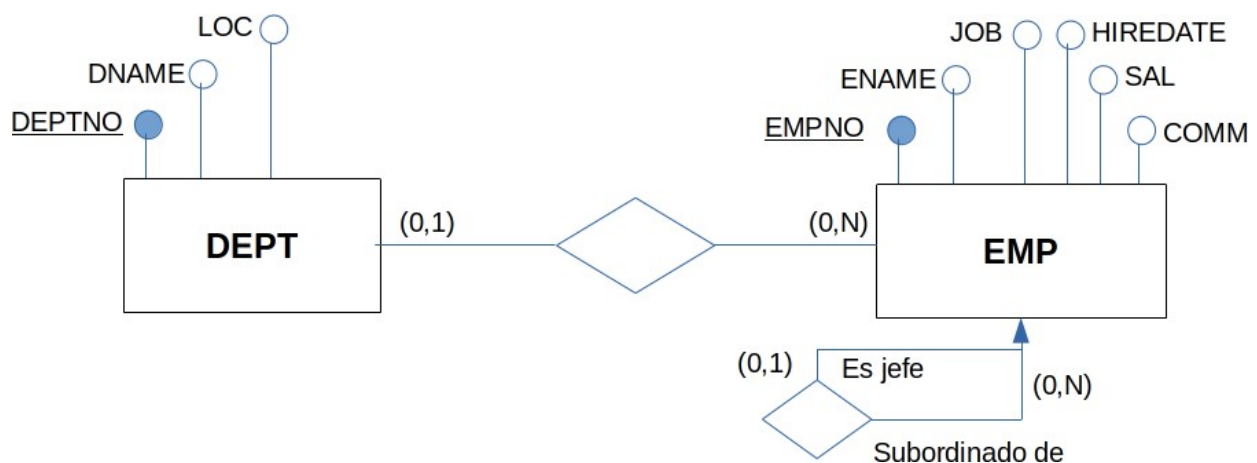
FASE 1: SOLUCIÓN NoSQL MONGODB

OBJETIVO

El objetivo de esta práctica es afianzar los conocimientos impartidos en la parte teórica de la asignatura sobre el uso de la solución NoSQL MongoDB.

TRABAJO A REALIZAR

A partir del siguiente esquema entidad-interrelación:



1. Proponga una estructura de colecciones de documentos MongoDB para almacenar la información.
2. Realizar una carga de información lo suficientemente completa para que le sirva como prueba y validación.
3. Realice varias consultas generando informes de salida que le permitan comprobar el acceso a la información.



Análisis, Diseño y Procesamiento de Datos Aplicados a las Ciencias y a las Tecnologías

PRÁCTICAS (APARTADO III)

FASE 2: APLICACIONES CIENTÍFICAS Y EMPRESARIALES (I)

OBJETIVO

El objetivo de esta práctica es afianzar los conocimientos impartidos en la parte teórica de la asignatura sobre la caracterización y predicción de actividades/propiedades de compuestos.

TRABAJO A REALIZAR

Utilizando las bases de datos de compuestos del ejemplo práctico desarrollado en la parte teórica de la asignatura, implementar el código python necesario para predecir (clasificación) el tipo de actividad biológica (Active/Inactive).

Recomendaciones para la implementación:

- 1- Utilice alguno de los modelos de clasificación implementados en la biblioteca scikit-learn.
- 2- El tipo de actividad de cada compuesto (variable respuesta) está almacenado en el campo “class” de la colección “molecules”.
- 3- Como variables predictoras se utilizará el fingerprint molecular de cada compuesto. Recuerde que en la base de datos sólo están almacenadas las posiciones de los bits con valor “1”.

Las variables predictoras y la respuesta quedarían representadas de siguiente forma:

id_Compuesto	FP1	FP2	...	FP1024	class
comp-1	0	0		1	Active (1)
comp-2	1	0		1	Inactive (0)
comp-N	1	0		0	Active (1)

- 4- Las columnas donde todos los valores sean ‘0’ pueden ser eliminadas.
- 5- Para evaluar el desempeño de los clasificadores se deben utilizar métricas clásicas como G-Mean, Kappa, Accuracy, AURoc, etc.



Análisis, Diseño y Procesamiento de Datos Aplicados a las Ciencias y a las Tecnologías

PRÁCTICAS (APARTADO III)

ENTREGA Y REQUISITOS

- El trabajo no tendrá una limitación en cuanto a volumen/tamaño, pero si en duración, debiendo realizarse en el plazo prefijado.
- La memoria principal del trabajo se entregará en formato pdf, siendo de obligatorio cumplimiento la entrega de todos los ficheros fuente utilizados (de código, .doc, figuras, imágenes, etc.).
- Para la entrega se utilizará la tarea moodle que será creada antes de la fecha límite.
- La primera página será una portada, con indicación de: titulación, asignatura, datos del alumnos, fecha y título.
- Todos los ficheros (pdf de la memoria, código fuente, etc) se entregará en un único fichero con formato .zip con el siguiente formato: NOMBRE-PRIMER APELLIDO-SEGUNDO APELLIDO.zip