

PRIMER ENTREGA

SQL FLEX

VICTORIA MARQUEZ

SQL y meta análisis

El cambio climático es uno de los factores de perturbación antrópica más importantes que afectan actualmente a los ecosistemas. Tanto las plantas como sus polinizadores insectos se ven afectados por el cambio climático. Las alteraciones en las fenologías de plantas y polinizadores debido al cambio climático pueden provocar desajustes en sus interacciones, lo que a su vez puede comprometer los servicios de polinización y el funcionamiento de los ecosistemas. Uno de los objetivos de mi proyecto post doctoral es realizar un meta análisis para examinar los efectos del cambio climático sobre las fenologías de floración de plantas y de los polinizadores en estudios realizados en distintos ecosistemas y con diferentes enfoques metodológicos, con el fin de identificar patrones, sesgos metodológicos y posibles desajustes planta-polinizador a nivel global.

Un meta-análisis es una técnica estadística muy utilizada en ecología que tiene como objetivo sintetizar cuantitativamente los resultados de múltiples estudios independientes sobre un mismo tema. Su objetivo es identificar patrones generales, estimar efectos promedio y evaluar la consistencia de los resultados entre estudios. En lugar de basarse en una sola investigación, el meta-análisis reúne datos de muchos trabajos científicos previos, y mediante herramientas estadísticas permite responder preguntas con mayor poder y robustez. Es especialmente útil cuando los estudios individuales muestran resultados variables o contradictorios.

En las etapas iniciales de un meta-análisis, se generan bases de datos extensas, producto de la recopilación de información de muchos estudios. Estas bases pueden incluir cientos o miles de registros relacionados con especies, ubicaciones, variables fenológicas, parámetros climáticos, años de estudio, entre otros. Dado este volumen y complejidad, el uso de bases de datos relacionales y en particular el lenguaje SQL resulta sumamente útil para organizar, consultar, cruzar y analizar esta información de forma robusta, eficiente y reproducible. Por ejemplo, una base de datos relacional permite: relacionar especies con

sitios de muestreo y condiciones climáticas, clasificar estudios según enfoque metodológico, año o región y consultar con rapidez tendencias específicas o generar subconjuntos de datos para análisis más focalizados.

Propongo con este objetivo la creación de una base de datos que contenga la información recopilada a lo largo de los múltiples estudios independientes analizados la cual va a contener las siguientes entidades (tablas) y atributos:

Entidad	Descripción	Atributos	Relación conceptual
Estudio	Contiene información básica de cada trabajo incluido en la base de datos.	id_estudio INT, título VARCHAR(100), año YEAR, DOI VARCHAR(100), tipo de estudio VARCHAR(100)	Cada estudio está vinculado a uno o más sitios y a una o múltiples especies de plantas, polinizadores o ambos.
Planta	Almacena datos sobre las especies vegetales estudiadas	id_planta INT, nombre científico VARCHAR(150), familia VARCHAR(100), forma de vida VARCHAR(100), id_estudio INT	Son polinizadas por una o más especies de polinizadores

Polinizador	Almacena datos sobre los polinizadores registrados en los estudios.	id_polinizador INT , nombre científico VARCHAR(100), grupo funcional VARCHAR(100), id_estudio INT	Polinizan una o más plantas a través de
Interacción	Almacena datos sobre las interacciones entre plantas y polinizadores observadas en los estudios.	id_interacción INT , id_planta INT , id_polinizador INT , id_sitio INT , id_estudio INT	Une especies de plantas y polinizadores en un sitio y fecha determinada.
Sitio	Registra los sitios donde se llevaron a cabo los estudios.	id_sitio INT , nombre del sitio VARCHAR(100), país VARCHAR(100), latitud DECIMAL(9,6), longitud DECIMAL(9,6), tipo de ecosistema VARCHAR(100), id_sitio INT	Lugar donde se registran las interacciones, la fenología y los datos de clima .

Fenología_planta	Registra la fenología de la floración de las plantas observadas.	id_feno_planta INT, inicio_floración DATE, pico_floracion DATE, fin_floracion DATE, id_planta INT, id_sitio INT	Relaciona plantas , sitios y fechas.
Fenología_polinizador	Registra la actividad fenológica de los polinizadores.	id_feno_po INTI, inicio_actividad DATE, pico _actividad DATE, fin_actividad DATE, id_polinizador INT, id_sitio INT	Relaciona polinizadores , sitios y fechas.
Clima	Contiene datos climáticos asociados a cada sitio de muestreo.	id_dato_clima INT, temperatura media FLOAT, precipitación FLOAT, id_sitio INT	Vinculados a cada sitio y pueden usarse para analizar su efecto en la fenología .

Tabla 1. Descripción de las tablas propuestas, sus atributos y relación conceptual con el resto de las tablas. Las claves primarias se encuentran en azul y las claves foráneas en rojo.

Diagrama de entidad de relaciones conceptual:

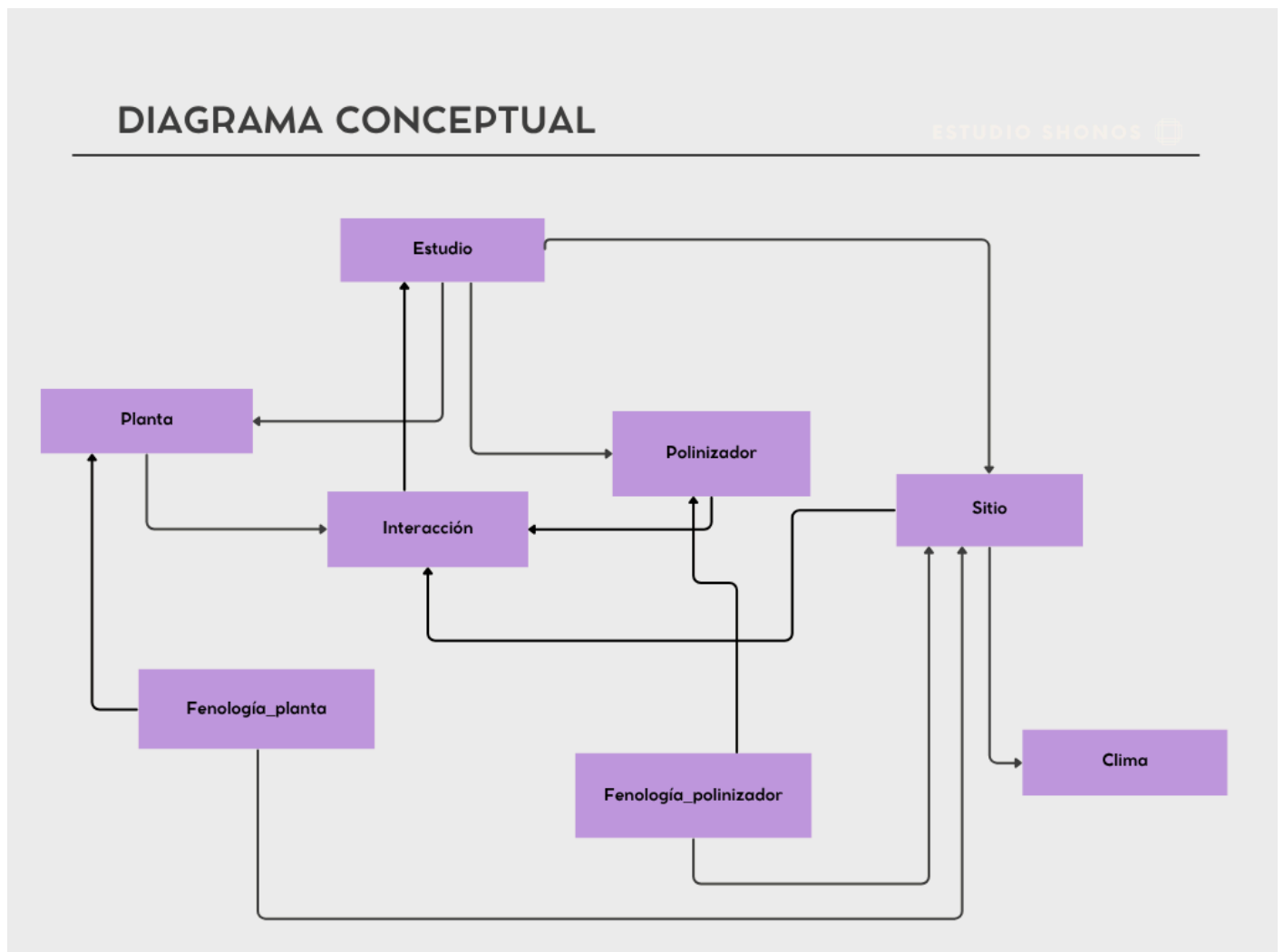


Fig.1 Diagrama de relaciones propuesta para la base de datos.

Importación de datos

En este caso, dada la especificidad y complejidad de los datos propuestos utilicé Python para crear archivos csv con valores aleatorios, pero relacionados entre sí. Luego realicé la importación de los datos en MySQL Workbench de la siguiente manera.

1. Seleccioné la tabla destino de los datos que voy a importar

2. Click derecho sobre el nombre de la misma y seleccione *Table Data Import Wizard*

3. En *File Path* elegí la carpeta de origen de mis archivos csv y el archivo en sí mismo

4. Seleccioné la opción *Use existing table* por que mis tablas ya están creadas

5. En la siguiente pestaña seleccioné el *Encoding* utf-8 y finalmente importo la información

Algo importante a chequear es que la identidad, número y nombre de las columnas en el archivo csv coincidan con las columnas de nuestra tabla en MySQL.

Creación de objetos

Vistas

- 1) vista_fenologia_mismatch: tiene como objetivo mostrar las fenológicas de las plantas y polinizadores que estén en lo mismos sitios. Es útil para detectar posibles desajustes y solapamientos fenológicos entre plantas y polinizadores. Las tablas involucradas son: fenologia_planta, planta, sitio, fenologia_polinizador y polinizador.
- 2) vista_interacciones_taxonomia: me permite ver todas las interacciones planta-polinizador junto a la información taxonómica (nombre científico de polinizador y planta). Las tablas involucradas son: planta, polinizador e interacciones.
- 3) vista_estudios_detalle: resume por estudio las especies vegetales y polinizadoras involucradas, junto con los sitios donde se realizaron. Sirve para hacer análisis bibliométricos que pueden ser importantes en meta-análisis. Las tablas involucradas son: estudio, planta, polinizador, sitio.

Funciones

- 1) duracion_floracion_planta (p_id_feno_planta INT): calcula la duración total en días de la floración de una planta a partir de su fecha de inicio y fin. Tabla involucrada: fenologia_planta
- 2) numero_interacciones_por_estudio(p_id_estudio INT): retorna la cantidad de interacciones registradas en un estudio específico. Tabla involucrada: interacciones

Procedimientos

- 1) sp_insertar_interaccion (p_id_planta, p_id_polinizador, p_id_estudio): inserta una nueva interacción entre una planta y un polinizador asociada a un estudio. Tabla involucrada: interacciones
- 2) ver_fenologia(p_nombre_cientifico): devuelve las fechas de floración (inicio, pico y fin) de una planta según su nombre científico. Permite consultar rápidamente la fenología de una especie vegetal específica. Tablas involucradas: planta y fenologia_planta.

Triggers

- 1) trg_validar_fechas_fenologia: verifica al insertar en fenologia_planta que la fecha de inicio no sea posterior a la de fin. Garantizar la integridad temporal de los registros fenológicos. Tablas involucradas: fenologia_planta. Trigger: BEFORE INSERT
- 2) trg_log_cambio_nombre_planta: registra en una tabla log cada vez que se modifica el nombre científico de una planta. Llevar un historial de cambios en nomenclatura científica, útil para trazabilidad de datos. Tablas involucradas: planta y log_cambios_nombre_cientifico. Trigger BEFORE UPDATE
- 3) trg_log_cambio_nombre_polinizador: igual que el anterior, pero para los nombres científicos de polinizadores. Tablas involucradas: polinizador y log_cambios_nombre_cientifico. Trigger BEFORE UPDATE