

ENTREGA 4: Persistencia

Objetivos de la entrega

- Incorporar nociones de persistencia de datos en un medio relacional.
- Incorporar nociones de la técnica de mapeo objeto – relacional.
- Incorporar nociones de desnormalizaciones del modelo relacional

Unidades del Programa Vinculadas

- Unidad 2: Herramientas de Concepción y Comunicación del Diseño
- Unidad 5: Diseño de Datos y Estrategias de Persistencia
- Unidad 6: Diseño de Arquitectura
- Unidad 8: Validación del Diseño

Alcance

- Persistencia del modelo de objetos previamente generado por Servicio
- Normalización de la información
- Servicio de Estadísticas sobre los hechos disponibles
- Soporte para incorporación de videos e imágenes
- Exportación de datos en formato CSV
- Soporte para búsqueda por texto libre

Dominio

Normalización de Hechos e Información

Una de las principales problemáticas que surgen al integrar servicios externos (como fuentes proxy o datasets provistos por terceros) es la inconsistencia en la representación de los datos. Las distintas fuentes de hechos pueden utilizar diferentes convenciones para nombres de categorías, ciudades, municipios, provincias, países, y formatos de fecha u otros atributos clave.

Por ejemplo, un mismo hecho podría estar categorizado como “Incendio Forestal” en una fuente, “fuego forestal” en otra; o una ubicación puede aparecer como “CABA” o “Ciudad Autónoma de Buenos Aires”, dependiendo del origen. Estas diferencias pueden hacer que los hechos recibidos de distintos orígenes parezcan diferentes, aunque en la realidad representan la misma pieza de información.

Por este motivo, antes de persistir los datos en un medio relacional, es fundamental aplicar un proceso de normalización que estandarice los valores clave. Este proceso de curaduría deberá contemplar la unificación de nombres y códigos de provincias, municipios, ciudades, mapeo de categorías, validar y convertir fechas y resolver ambigüedades entre los hechos.

Servicio de Estadísticas

Se desea desarrollar un servicio que, conectándose con el servicio agregador, genere periódicamente estadísticas sobre el uso del sistema y las muestre mediante una interfaz gráfica al usuario. Cada cierto intervalo de tiempo, se recalcularán las estadísticas y se actualizarán los *dashboards* de la interfaz.

Específicamente, se piden obtener datos que permitan responder las siguientes preguntas:

- De una colección, ¿en qué provincia se agrupan la mayor cantidad de hechos reportados?
- ¿Cuál es la categoría con mayor cantidad de hechos reportados?
- ¿En qué provincia se presenta la mayor cantidad de hechos de una cierta categoría?
- ¿A qué hora del día ocurren la mayor cantidad de hechos de una cierta categoría?
- ¿Cuántas solicitudes de eliminación son spam?

El Servicio de Estadísticas deberá soportar la exportación de los datos generados mediante formato CSV.

Requerimientos detallados

1. Se deberán persistir las entidades del modelo planteado. Para ello se debe utilizar un ORM.
2. Se deberá normalizar los valores claves referidos a los hechos.
3. Se deberá implementar el Servicio de Estadísticas.
4. Se deberá implementar la exportación de las estadísticas en formato CSV.
5. (Bonus) Se deberá implementar una estrategia de búsqueda de texto libre.

Entregables

1. **Modelo del Dominio:** actualización del modelo de Diagrama de Clases con las funcionalidades previstas en esta entrega.
2. **Justificaciones de Diseño:** Documento y Diagramas Complementarios.
3. **Modelo de datos:** Diagrama de entidad-relación físico por Servicio.
4. **Justificaciones y Consideraciones de Diseño Relacional.**
5. **Implementación** en código de los requerimientos de la presente entrega.
6. **Documentación de API** del Servicio de Estadísticas.
7. (Bonus) **Implementación** de alguna herramienta que permita búsqueda por texto libre.¹

Observaciones

Cada servicio debe tener persistencia es un esquema relacional propio. Se puede compartir un mismo motor a fines de simplicidad.

¹ Consultar [MySQL Full Text Search](#) o [Elastic Search](#) para mayor referencia.