

Proyecto 2 Data Repository

1. Descripción

Los Repositorios de Datos (Data Repositories) son repositorios científicos donde se pueden poner en acceso abierto conjuntos de datos asociados a proyectos de investigación.

Además, los Repositorios de Datos brindan transparencia y rendición de cuentas, al visualizar los resultados de los proyectos de investigación financiados con fondos públicos y privados. Estos resultados no solo incluyen los conjuntos de datos, sino también, publicaciones relacionadas y la información de los investigadores participantes.

El TEC ha decidido implementar su propio Repositorio de Datos, y usted ha sido seleccionado para realizar dicha implementación en una plataforma web para tener acceso 24/7.

2. ¿Qué se busca con esta tarea?

- Construir un modelo Entidad-Relación de una aplicación.
- Implementar aplicaciones de software a partir de requerimientos de software.
- Fortalecer las habilidades investigativas y de redacción científica.
- Practicar la experimentación y la resolución de problemas (divide y vencerás).
- Ejercitar la toma de decisiones.
- Fomentar el trabajo cooperativo entre los miembros del equipo.
- Fortalecer las habilidades blandas del estudiante.
- Implementar aplicaciones en PHP.
- Utilizar el motor de bases de datos MySQL.

3. Contexto y las funcionalidades esperadas

Se deberá implementar en PHP usando como motor de bases de datos MySQL una aplicación web que funcione como el Repositorio de Datos del TEC. El sistema debe permitir gestionar la información de diferentes entes, los cuales se describen a continuación, obtenidos a partir de la definición de requerimientos realizada. Además, se cuenta con una vista pública, donde se dispone de acceso abierto a toda la información del Repositorio.

3.1. Módulo de Administración

3.1.1. Investigadores

El sistema debe gestionar la información de los investigadores (CRUD). Cada investigador se identifica con su número ORCID¹. Para registrar un nuevo investigador se necesita: su ORCID, su nombre, sus apellidos (puede ser solo uno o dos), su información de contacto (dirección completa, teléfono de oficina, correo electrónico), afiliación actual, áreas de interés (cualquier cantidad) y foto. Ejemplo: <https://scholars.duke.edu/person/a.aballay>

3.1.2. Financiamientos

El sistema debe gestionar la información de las fuentes de financiamiento (CRUD). Cada fuente se identifica con un identificador alfanumérico único. Para registrar una nueva fuente de financiamiento se necesita: su identificador, nombre, organismo que otorga el financiamiento, organismo que lo administra, investigador a cargo, fecha de inicio (año), fecha de finalización (año), estado (activo, finalizado, pendiente), tipo de financiamiento (público o privado) y monto (en dólares). Un financiamiento en estado finalizado no puede ser modificado/borrado.

3.1.3. Proyectos

El sistema debe gestionar la información de proyectos de investigación (CRUD). Cada proyecto se identifica con un identificador alfanumérico único. Para registrar un nuevo proyecto de financiamiento se necesita: su identificador, nombre, descripción, investigadores participantes (uno o más), fecha de inicio (año), fecha de finalización (año), estado (activo, finalizado, pendiente). Un proyecto en estado finalizado no puede ser modificado/borrado. Los proyectos siempre están asociados a fuentes de financiamiento.

3.1.4. Publicaciones

El sistema debe gestionar la información de publicaciones científicas (CRUD). Cada publicación se identifica por su DOI². Para registrar una nueva publicación se necesita: su identificador DOI, título, autores (investigadores), resumen (abstract), palabras clave (cualquier cantidad), mes y año de publicación, fuente, y documento (únicamente formato PDF). Las publicaciones siempre están asociadas a proyectos. Ejemplo: <https://repositoriotec.tec.ac.cr/handle/2238/6706>, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.385.4383&rank=2>

3.1.5. Datasets (conjuntos de datos)

El sistema debe gestionar la información de datasets (CRUD). Cada dataset se identifica por su nombre. Para registrar un nuevo dataset se necesita: su nombre, fecha de creación, fecha de recolección (año inicio – año final), descripción, investigador de contacto, archivo (cualquier formato/tipo), análisis y conclusiones (un solo campo). Ejemplo: <https://repositoriotec.tec.ac.cr/handle/2238/6706>, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.385.4383&rank=2>

¹ <https://orcid.org/>

² <https://www.doi.org/>

3.1.6. Asociaciones

- Un financiamiento puede financiar cualquier cantidad de proyectos.
- Un proyecto puede tener cualquier cantidad de fuentes de financiamiento.
- Las publicaciones pueden estar asociadas a un proyecto como máximo, y a cualquier cantidad de datasets.
- Los datasets pueden estar asociados a cualquier cantidad de proyectos y/o de publicaciones.

3.1.7. Archivos

Se deben almacenar en la base de datos los archivos correspondientes a las fotos de los investigadores, las publicaciones en PDF, y los datasets. Los datasets pueden ser archivos grandes (varios gigabites). Las publicaciones y los datasets deben ser descargables.

3.1.8. Usuarios administradores

Deben existir usuarios administradores que usan el módulo de administración. Para estos usuarios se debe guardar el correo electrónico y la contraseña (encriptada). Debe existir un usuario administrador inicial. Un usuario administrador puede agregar nuevos usuarios administradores. El acceso al módulo de administración se hace desde la vista pública, con un enlace tipo 'acceso para administradores' en una parte discreta.

3.2. Vista pública

La interfaz principal del Repositorio de Datos será una interfaz pública de acceso abierto (cualquier usuario puede navegar sin necesidad de estar registrado).

3.2.1. Navegabilidad

Esta interfaz deberá permitir navegar por:

1. Investigadores: Deberá desplegar una lista de todos los investigadores (nombre completo), con opción de ordenamiento. Se puede hacer clic sobre cualquier investigador, y se debe cargar toda su información, junto con la lista de proyectos (enlace al detalle del proyecto), publicaciones (enlace al detalle de la publicación), lista de datasets (enlace al detalle del dataset). Ejemplo: <https://scholars.duke.edu/people>
2. Proyectos: Deberá desplegar una lista de todos los proyectos (nombre), con opción de ordenamiento. Se puede hacer clic sobre cualquier proyecto, y se debe cargar toda su información, junto con la lista de publicaciones asociadas (enlace al detalle de la publicación), fuentes de financiamiento (enlace al detalle de la fuente: incluye lista de proyectos asociados) y lista de datasets asociados (enlace al detalle del dataset).
3. Publicaciones: Deberá desplegar una lista de todas las publicaciones (nombre), con opción de ordenamiento (nombre y año). Se puede hacer clic sobre cualquier publicación, y se debe cargar toda su información, junto con la lista de proyectos asociados (enlace al detalle del proyecto), lista de investigadores (enlace al detalle del investigador) y lista de datasets asociados (enlace al detalle del dataset). Se debe mostrar la cantidad de veces que se ha descargado la publicación (archivo pdf).
4. Datasets: Deberá desplegar una lista de todos los datasets (nombre), con opción de ordenamiento (nombre y año). Se puede hacer clic sobre cualquier dataset, y se debe

cargar toda su información, junto con la lista de proyectos asociados (enlace al detalle del proyecto), y lista de publicaciones asociadas (enlace al detalle de la publicación). Se debe mostrar la cantidad de veces que se ha descargado el dataset.

3.2.2. Búsqueda

La interfaz principal debe tener una barra de búsqueda. En la búsqueda el usuario introduce un conjunto de palabras, y se encuentran todas las fuentes de financiamiento, proyectos, publicaciones o datasets que coincide en su nombre, descripción, resumen o palabras claves. Se debe mostrar la lista de resultados, donde se indica el nombre, el tipo de resultado y el enlace al detalle. El usuario puede seleccionar el tipo de resultado que desea (por ejemplo, solo buscar publicaciones, o datasets, o ambos).

3.2.2.1. Estadísticas

En esta sección se deben poder seleccionar y visualizar las siguientes estadísticas:

1. Cantidad de financiamientos: filtrados por su fecha de inicio, fecha de finalización, estado, tipo de financiamiento.
2. Cantidad de proyectos: filtrados por su fecha de inicio, fecha de finalización, estado, tipo de financiamiento.
3. Cantidad de publicaciones: filtradas por año.
4. Cantidad de datasets: filtrados por año.
5. Cantidad de descargas totales de publicaciones.
6. Cantidad de descargas totales de datasets.
7. Top 5 de publicaciones más descargadas: filtrados por período de tiempo (año inicio, año fin).
8. Top 5 de datasets más descargados: filtrados por período de tiempo (año inicio, año fin).
9. Top 5 de investigadores con más publicaciones.
10. Top 5 de proyectos con más publicaciones.
11. Top 5 de proyectos con más datasets.

IMPORTANTE: todos los filtros son opcionales (el usuario escoge si los puede aplicar o no).

4. Diseño

Deben tomar todas las decisiones de diseño pertinentes: tablas, restricciones, etc. Debe procurar que el esquema de la base de datos sea lo más robusta posible, donde se incluyan todas las restricciones necesarias (dominios, restricciones de borrado por llaves foráneas, etc.) para que la base de datos siempre se mantenga en un estado válido. El diseño de la base de datos debe ser eficiente en cuanto a espacio y rapidez de acceso a los datos.

5. Aspectos técnicos

El proyecto deberá estar escrito en el lenguaje de programación PHP con conexión a MySQL, correr sobre el servidor web Apache y deberá de funcionar en el sistema Linux. Puede utilizar aplicaciones como ampps³ que tienen un ambiente de desarrollo integrado.

En la documentación se debe incluir las librerías externas utilizadas, **y todos los requisitos y pasos necesarios** para la ejecución de la aplicación web.

6. Documentación

La documentación es un aspecto de gran importancia en el desarrollo de programas, especialmente en tareas relacionadas con el mantenimiento de los mismos.

Para la documentación interna, deberán incluir comentarios descriptivos para las funciones principales de cada clase, con sus entradas, salidas, y restricciones.

La documentación externa deberá incluir (un solo archivo en formato PDF y en el orden especificado):

1. Portada.
2. Manual de usuario: Debe indicar como correr el programa paso a paso (el archivo y como correrlo), cuales funcionalidades están disponibles y cuál es su funcionamiento. Debe incluir los requisitos para correr el programa. Debe guiar al usuario para poder utilizar el programa exitosamente. Haga uso de **screenshots** y comentarios.
3. Pruebas de funcionalidad: Debe demostrar evidencia de todas las funcionalidades implementadas en el proyecto programado. Para esto realice la ejecución del proyecto y tome **screenshots** dejando evidencia del resultado de la ejecución, según los requerimientos esperados. Debe mostrar evidencia también de la robustez de la aplicación (por ejemplo, qué pasa cuando el usuario ingresa un dato equivocado).
4. Diagramas: Debe generar los diagramas ER para la base de datos del sistema. Puede utilizar la notación de Chen, notación Crow's Foot o notación UML. Recuerde incluir las tablas intermedias para el funcionamiento del proyecto. Además, debe incluir el diseño lógico de la base de datos.
5. Diccionario de datos: Debe dar un diccionario de datos completo. Por cada tabla debe indicar el propósito de la tabla, y todos sus campos (incluye nombre, tamaño, tipo de datos, descripción y justificación de porque se eligió ese tipo de datos y tamaño).
6. Análisis de resultados: objetivos alcanzados, objetivos no alcanzados, y razones por las cuales no se alcanzaron los objetivos (en caso de haberlos). Además, se debe incluir una tabla donde se indique la distribución del trabajo realizado por cada uno de los integrantes del equipo.

³ <http://www.ampps.com/>

7. Evaluación

1. Documentación externa 25%:
 - a. Diagramas y diccionario de datos: 15%
 - b. Manual de Usuario, pruebas de funcionalidad, y análisis de resultados: 10%.
2. Módulo administrativo: 40%
3. Vista pública: 35%
4. EXTRA: Creación de usuarios públicos, además los usuarios pueden guardar “recursos” favoritos, los cuales pueden visualizar/editar en cualquier momento (10%).

8. Aspectos administrativos

1. Debe crear un archivo .zip (“P2.zip”) que contenga únicamente un archivo **info.txt** y 2 carpetas llamadas **documentacion** (con el archivo de la documentación) y **solucion_computacional**, en la primera deberá incluir el documento PDF solicitado y en la segunda los archivos y/o carpetas necesarias para la implementación de este proyecto. El archivo **info.txt** debe contener la siguiente información (calidades):
 - a. Nombre del curso
 - b. Número de semestre y año lectivo
 - c. Nombre del Estudiante x 3
 - d. Número de carnet x 3
 - e. Número de tarea programada
 - f. Fecha de entrega
 - g. Estatus de la entrega (definido por el responsable de la implementación de la tarea, debe ser **CONGRUENTE** con la solución entregada):
[Deplorable | Regular | Buena | MuyBuena | Excelente | Superior]
2. Deberá subir el archivo antes mencionado al TEC Digital en el curso de BASES DE DATOS GR 01, en la asignación llamada “Proyecto Programado 2” debajo del rubro de “Proyectos programados”. Se deberá subir el archivo antes de las 11:50 PM del domingo 29 de octubre del 2017.
3. Dentro de la carpeta de **solucion_computacional**, deberá incluir un archivo .txt (**Nombre1.Nombre2.Nombre3.txt**) que contenga todo el texto de la solución del o de los archivos presentados (las implementaciones). Este archivo puede ser revisado en el sistema de Control de Plagio del TEC Digital. **Todo el código de cada proyecto debe ser 100% original, y no se va a tolerar el plagio.** Este archivo también deberá ser subido al TEC Digital en el curso de BASES DE DATOS GR 01, en la asignación llamada “Proyecto Programado 2 (archivo TXT)” debajo del rubro de “Proyectos programados”. Se deberá subir el archivo en el mismo tiempo que el archivo ZIP.
4. Las tareas deberán ser revisadas con el profesor o el asistente. Todos los miembros del grupo deberán participar de la revisión, de lo contrario a la nota obtenida grupalmente se le descontará 15%. Además, si durante la revisión el estudiante no puede demostrar conocimiento de la solución presentada, se le descontará hasta un 15% de la nota grupal.
5. La tarea se hará en grupos de 3 personas.

6. La tarea vale un 20% de la nota del curso.
7. Antes de la entrega del proyecto se pasará en clase una lista donde deben anotar los miembros de su equipo de trabajo. Opcionalmente, pueden mandar un correo electrónico con el asunto “[BD 217] P2:GRUPO” e indicar los nombres de las personas que conforman el grupo de trabajo para esta tarea programada. Esta información es necesaria para crear los grupos en el TEC Digital. Si no se anotan en la lista, o no envían el correo, se calificará la tarea de forma individual al miembro del equipo que la suba.

IMPORTANTE: CONOCIMIENTO DE LA SOLUCIÓN PRESENTADA. En la revisión del trabajo, los estudiantes deben demostrar un completo dominio de la solución implementada, tanto desde el punto de vista técnico (uso de herramientas) como de la funcionalidad del proyecto. La revisión se puede hacer individualmente o en grupos, examinando la solución o temas específicos aplicados en el proyecto. Todos los integrantes del grupo deben tener el mismo conocimiento de la solución presentada.

IMPORTANTE: DIAGRAMAS. Se recomienda que al realizar el diseño conceptual de la base de datos y el diseño lógico de la base de datos estos sean consultados con el profesor para su revisión (puede utilizar el horario de consulta o coordinar otros espacios).