

Proyecto - Parte 2 - Asignación de Programación Bases de Datos Textuales / Documentales

PRESENTACIÓN

La asignación consiste en crear una aplicación en Oracle Application Express (APEX) que permita almacenar, catalogar y consultar - usando SQL y operadores textuales - documentos de diversos formatos, incluyendo PDF, Word, texto, HTML, PowerPoint, etc. La aplicación a desarrollar debe cumplir con todos los requerimientos que se plantean en este documento. Esto por requerirá de crear un índice textual y de construir la aplicación en APEX utilizando SQL con operadores textuales, al mismo tiempo que utilizar las funciones de Oracle Text en PL/SQL.

Se recomienda la consulta y lectura del manual de “Oracle Text Application Developer's Guide”, que se encuentra en el Oracle Technology Network (OTN), [aquí](#). Adicionalmente, también cuenta con el manual Oracle Text Reference, que está disponible [aquí](#) y que contiene la descripción de todos los comandos/operadores textuales en la base de datos de Oracle.

Para familiarizarse con el desarrollo en Oracle Application Express cuenta con diversos manuales en OTN. Para ello, vaya a la sección de “Application Development” de la documentación de la base de datos Oracle ([aquí](#)) y haga scroll hacia abajo hasta encontrar la sección de Application Express. Un manual muy útil es el “Application Express Application Builder User's Guide”.

Si requiere agregar algún elemento al contexto o asumir algún aspecto, indíquelo claramente junto con la entrega del script en Moodle y justifique por qué lo está agregando o asumiendo. La justificación para asumir cualquier cosa debe ser válida y se evalúa como parte de la asignación. Asumir algo para super-simplificar el escenario no es admisible. Si tiene alguna duda, consúltelo con el profesor.

CONFIGURACIÓN INICIAL

Para iniciar el proyecto, debe seguir la Guía de Configuración Inicial que se provee en conjunto con este documento.

REQUERIMIENTOS GENERALES

La aplicación a construir debe contemplar todos los puntos de función que se describen a continuación. Como referencia práctica, una implementación sugerida de la aplicación solicitada se puede acceder a través del siguiente enlace (usuario: demo/ password: demo1234)

<https://apex.oracle.com/pls/apex/f?p=72862>

La aplicación en APEX debe contemplar los siguientes aspectos:

1. Permitir cargar documentos en diversos formatos (word, pdf, powerpoint, html, texto, keynote, pages, numbers, excel) en una tabla en la base de datos, específicamente en una columna tipo Binary Large Object (BLOB).
2. Proveer una interfaz para que el usuario pueda cargar los documentos, completar la información del documento y darle mantenimiento a esta información, así como para eliminar documentos de la base de datos.

3. Todos los documentos deben ser cargados al servidor de base de datos a partir de archivos de la máquina local.
4. Los documentos deben tener, además del archivo propiamente dicho en el BLOB, los siguientes elementos de datos. Usted es libre de crear cualquier otra columna que considere necesaria.

Titulo, 500 caracteres

Fecha de creación

Notas: 1000 caracteres

Tipo de Documento: Artículo, Brochure, Obra Literaria, Ensayo, Presentación, Manual, Documento Técnico y Otro.

Tipo de Archivo: que debe ser "calculado" automáticamente a partir del archivo que se carga. Considere tener PDF, Word, Excel, PowerPoint, Texto, HTML, Keynote, Pages, Numbers, y para todos los demás tipos simplemente indicar "Otro".

Disponible: Si o No

Autores: 300 caracteres

Fecha de Carga, que debe contener la fecha en la que se cargó el archivo

Tamaño del Archivo: que debe ser calculado a partir del archivo que se carga

Usuario que carga: username del usuario que cargó el archivo.

5. Debe proveerse una interfaz para que los usuarios puedan obtener un listado de los archivos existentes, y descargarlos u actualizar sus datos, así como agregar archivos en al base de datos.
6. Debe proveerse una página de búsqueda, en la que los usuarios puedan buscar los archivos a partir de una combinación de al menos los siguientes campos:
 - i. Texto Libre: en el contenido del documento, utilizando operadores textuales de Oracle Text (y el operador CONTAINS en el SELECT).
 - ii. Score mínimo: que debe ser un valor seleccionable de una lista, con las siguientes posibilidades. > 90, > 80, > 75, > 50, > 25. El Score, se refiere a la pseudo-columna SCORE generada por una consulta textual en Oracle Text (usando CONTAINS).
 - iii. Tipo de Archivo, para filtrar para un tipo particular seleccionado por el usuario
 - iv. Tipo de Documento, similar al anterior
7. Si un documento se marca como Disponible = No, entonces no debe aparecer en el resultado de la búsqueda.
8. El reporte de los documentos resultantes tras la búsqueda debe incluir los enlaces para:
 - a) descargar el archivo en el formato original
 - b) obtener una versión del archivo en formato HTML con los términos de búsqueda resaltados (en rojo y en negrita) en el documento
 - c) cada término resaltado en la versión en HTML debe incluir dos enlaces para navegar a la siguiente y a la anterior ocurrencia de la palabra en el archivo
 - d) obtener la lista de temas (THEMES) que el documento tiene, generados utilizando Oracle Text. Para clasificar los documentos, utilice el método unsupervised
 - e) obtener un resumen (GIST) del documento en general, o por tema de acuerdo a la selección del usuario (solo para los documentos en inglés o francés). En resumen (GIST) del documento debe generarse a nivel de SENTENCE

9. El reporte de los documentos resultantes debe indicar el total de documentos que cumplen con el criterio del texto libre de la búsqueda (operador textual). Nota: debe proveerse el numero de documentos que cumple el operador textual, sin considerar otros filtros que se hayan aplicado.
10. Debe proveerse un mecanismo para que el usuario puede re-sincronizar el indice textual, por ejemplo después de que ha cargado un conjunto nuevo de documentos.
11. La aplicación debe mantener una clasificación de los temas que los documentos tienen, así como el peso de ese tema para el documento. Esta clasificación ocurre manualmente en dos momentos: a) cuando el usuario selecciona ver los temas de un documento particular, b) cuando el usuario elige re-clasificar todos los documentos en la base de datos.
12. Deben incluirse gráficos que muestren:
 - a) La cantidad de documentos por tipo de documento,
 - b) La cantidad de archivos por tipo de archivo,
 - c) La cantidad de documentos por tema, para los 10 temas con más documentos.
13. Todas las condiciones de error deben ser manejadas para indicar al usuario el mensaje correspondiente.

EVALUACIÓN

Para completar la asignación, usted deberá subir a Moodle, antes de la fecha y al hora indicada en la definición del Proyecto, un export de la aplicación que construyó cumpliendo con los todos puntos de función que se han descrito. Si como parte de este export no se incluyen los “supporting objects”, entonces debe proveer un script en SQL con el que se creen los objetos (tablas, vistas, procedimientos almacenados, packages, triggers, etc) necesarios para que la aplicación funcione.

Debe incluir también cualquier dato inicial que sea necesario en sus tablas. No hace falta que se incluyan documentos en las tablas; el profesor cargará los suyos propios en la aplicación para comprobar que la aplicación cumple con los puntos de función indicados.

Para generar el export de la aplicación en APEX vaya al Application Builder seleccione la aplicación que usted creo y luego el icono "Export/Import". Luego seleccione "Export" y seleccione "Next". En la página que se encuentra ahora, seleccione "Export Application". Se iniciará la descarga de un archivo .sql (podría llamarse algo como **F101.sql**). Este es el archivo de la aplicación que debe subir a Moddle.

Si alguno de los archivos que usted entrega tiene errores al ser ejecutado en un esquema nuevo, no será posible revisar su tarea y por ende se afectará negativamente la calificación de su trabajo. Por lo tanto se sugiere fuertemente pruebe los archivos que entregue antes de hacerlo para asegurarse de que funcionan.