

Investigación 07: árboles de archivos

-Compresión y encriptación de datos por Huffman

Este es un algoritmo de compresión que se utiliza para comprimir datos con pérdida, lo que significa que la información original se pierde durante el proceso de compresión. Tal compresión es muy eficiente y puede reducir el tamaño de los datos en un porcentaje significativo sin perder demasiada información.

El algoritmo de compresión se basa en la frecuencia de aparición de cada carácter en el archivo que se está comprimiendo. Cada carácter se representa por un código de bits único que es más corto para los caracteres frecuentes y más largo para los menos frecuentes. De esta manera, los caracteres más frecuentes se comprimen más eficientemente reduciendo así el tamaño/peso del archivo.

Deforma un poco más técnica y detallada...

la técnica consiste en asignar a cada byte del archivo que se busca comprimir, un código binario compuesto por una cantidad de bits tan corta como sea posible. Esta cantidad será variable y dependerá de la probabilidad de ocurrencia del byte. En cambio los más extensos se utilizan para redefinir los bytes menos frecuentes.

Diccionario de datos

Este es un tipo de metadato que enumera de manera organizada los nombres, definiciones y características de cada uno de los campos o atributos de una base o conjunto de datos.

Estos diccionarios permiten entender e interpretar un conjunto de datos o base de datos, proporcionando información básica sobre los campos o variables que contiene. En otras palabras, pueden servir para:

- Saber que significa cada campo o variable
- Qué tipo de datos contiene
- Qué valores puede tomar o si usa un catálogo entre otros...

Es importante conocer sobre la existencia de los diccionarios de datos, pero de esta manera se pueden evitar malas interpretaciones y mal uso de los datos. Se pueden dar casos que por culpa de no existir un diccionario de datos, los mismos son inutilizables ya que su comparación se vuelve imposible.

Traducción y empaquetamiento

Primero, la traducción se refiere al proceso de convertir el código escrito en un lenguaje de programación a un código objeto ejecutable en una máquina específica, tal proceso es realizado por un intérprete o compilador.

La traducción entonces se da en dos posibles partes

- **Compilación:** Esta implica la conversión del código fuente a código objeto antes de su ejecución.
- **Interpretación:** Esta implica conversión y ejecución línea a línea. Estas dos acciones, a diferencia de la compilación, se hacen a la vez.

Empaquetamiento

Esta es, según la definición explícita; una de las capas del servicio web que permite establecer un formato de mensajes para comunicar aplicaciones, independientemente del lenguaje de programación o sistema operativo.

Esto es entonces, el proceso de crear un paquete o conjunto de archivos que contienen un programa o aplicación junto con cualquier otro archivo o recurso necesario para su ejecución.

Los paquetes pueden incluir bibliotecas de software, archivos de configuración, recursos multimedia y otros elementos necesarios para que el programa funcione correctamente.

- El empaquetamiento es utilizado por ejemplo en el desarrollo de software para crear archivos de instalación que pueden ser ~~instalados~~ distribuidos a los finales. Estos archivos suelen incluir un instalador que se encarga de instalar el programa en el sistema del usuario y configurar los elementos necesarios.

Árboles binarios de búsqueda con relación a la gestión de archivos

En este contexto, los árboles son utilizados para organizar y acceder a los archivos y directorios de manera muy eficiente, por ejemplo, en un sistema operativo, se puede representar la estructura de directorios y archivos del sistema. Cada nodo en ese caso representaría un directorio o archivo y los subárboles representarían los subdirectorios y subarchivos contenidos dentro del nodo.

Por la propia organización de los árboles, acceder a los archivos y directorios es una manera muy eficiente y rápida. Por ejemplo, cuando se busca un archivo en el sistema de archivos, se puede recorrer el árbol para encontrar el archivo deseado. Si está bien organizado, el tiempo de búsqueda se reduce significativamente.

Estos árboles se pueden buscar y actualizar mediante las operaciones básicas del propio árbol.

Carga de índices

Ya que un árbol consta de nodos con o sin hijos, este logra representar por cada nodo, un bloque de datos del disco.

La carga entonces, se basa en crear una estructura de árbol binario a partir de los datos almacenados en los bloques de datos del árbol de archivos. Así el árbol binario se utiliza para indexar los datos almacenados en el árbol de archivos, permitiendo así un acceso más eficiente a los datos.

Más a detalle, esta carga se realiza en varias etapas. En la primera, se lee cada bloque de datos del árbol de archivos y se procesa para extraer la clave y el puntero al siguiente bloque de datos.

En la segunda etapa, se construye el árbol a partir de las llaves extraídas, para así en la tercer etapa se guarda el árbol binario en el disco.

De esta manera, en lugar de usar el árbol de archivos se opera con el árbol pero binario de búsqueda.

Operaciones de inserción y búsqueda

- **Inserción:** Esta implica agregar un nuevo archivo o directorio en el árbol de archivos. Para realizar tal operación, se debe seguir el camino correcto desde la raíz del árbol hasta la ubicación donde se desea agregar el nuevo archivo o directorio. Luego, se crea un nodo nuevo en la ubicación deseada y se establecen las correspondientes referencias para los nodos padre e hijos.

- **Búsqueda:** Esta tiene como fin el encontrar un archivo o directorio en el árbol de archivos. Para ello, de la misma manera que en la inserción, solo se sigue el camino correcto desde la raíz hasta la ubicación deseada. Si se encuentra la el archivo o directorio, se regresa el mismo, sino se devuelve un mensaje de error.

Con estas dos anteriores operaciones, pueden derivarse gran parte de otras operaciones muy útiles con las que se puede operar el árbol (eliminar, modificar, etc)

Por ejemplo para eliminar, solo basta utilizar la operación de buscar. Igualmente para modificar pues realmente la operación de búsqueda es la que realmente trabaja buscando lo indispensable para las operaciones que la llaman (la ubicación del dato). Prácticamente si encuentra el dato, registra la ubicación del mismo y con la ubicación se pueden realizar las modificaciones que se deseen al dato.

Debido a la facilidad con la que se pueden modificar los datos existen y se recomiendan usar los diferentes mecanismos de protección de datos y seguridad para prevenir errores y pérdida de información.