

ENUNCIADO PARA LA PRÁCTICA 1

Introducción

La búsqueda es uno de los problemas clásicos, cuya importancia en un contexto como el actual, con voluminosas cantidades de información disponibles, refuerza su relevancia. Existen diversos algoritmos de búsqueda, y la conveniencia de utilizar unos u otros depende de diversos factores, como el hecho de que la colección en la que se busca esté ordenada o no, el tamaño de esta colección, de características del elemento que se busca, o la propia frecuencia de las búsquedas.

Objetivo

El objetivo de la práctica es comparar dos algoritmos de búsqueda y comprobar cuándo es más conveniente utilizar uno u otro, comparando la eficiencia en el **caso promedio**, usando una estimación basada en datos experimentales. Se comparará el algoritmo de *búsqueda secuencial* con el algoritmo de *búsqueda binaria*. Una vez hecha la estimación se determinará, si es posible, la curva que más se aproxima a los datos obtenidos. Por último se analizará qué algoritmo es más eficiente, indicando los **valores umbrales de tamaño del vector, n** , a partir de los cuales es aconsejable optar por uno u otro mecanismo de búsqueda. Esto se hará para el supuesto en que el vector está ordenado. Se tomarán en cuenta únicamente las operaciones y tiempos de las búsquedas, obviando el coste de poblar los vectores.

Una vez se haya realizado esta fase se hará una nueva comparación, partiendo de vectores no ordenados, cuyo objetivo será decidir cuándo interesa ordenar el vector para hacer una búsqueda binaria, o por el contrario, interesa realizar una búsqueda secuencial. En este caso se tomarán en cuenta los costes de ordenar el vector sobre el que se buscará. Para esta ordenación se tomará como base el método de *ordenación por inserción*.

Desarrollo de la práctica

La práctica constará de dos tipos de análisis: de las operaciones elementales y del tiempo empleado. En todos los casos hay que construir un programa para obtener los datos, en el cual se incluirán los subprogramas que implementan cada uno de los algoritmos.

- Se crearán y asignarán aleatoriamente vectores de distinto tamaño. Se probarán vectores de tamaño 10^i , con $i=1, 2, 3...$ hasta alcanzar el máximo valor que admita nuestro procesador. Es decir, se probarán tamaños $n= 10, 100, 1000...$ Para cada tamaño se llamará un cierto número de veces (por ejemplo, 30) al subprograma. Debe tenerse en cuenta que antes de cada llamada al subprograma se deberá volver a asignar valores aleatorios al vector.

- El programa servirá para obtener los datos necesarios para analizar los algoritmos. Proporcionará como salidas las tablas que indiquen:
 - El número de operaciones elementales promedio según el tamaño del vector. Se considerarán solamente las operaciones **entre elementos del vector** (no las comparaciones entre índices, por ejemplo).
 - El tiempo empleado según el tamaño del vector.

Usando estos datos se construirán dos tipos de gráficas, una para el número de operaciones elementales (comparaciones y/o movimientos), y otra para el tiempo empleado. A partir de ellas se intentará obtener una fórmula que las describa en función del tamaño del vector. Por ejemplo, se pueden ajustar las estimaciones a una curva polinómica, logarítmica, etc.

Para analizar los datos se sugiere utilizar programas que permitan representar gráficamente los datos obtenidos, por ejemplo LibreOffice Calc, Microsoft Excel o cualquier herramienta conocida que sea útil para este tipo de tareas.

Contar operaciones elementales

Se añadirán las instrucciones necesarias en el programa para contar el número de operaciones (comparaciones) en las que intervengan elementos del vector, en función del tamaño del mismo, y calcular los promedios.

Medir el tiempo empleado

La medición del tiempo empleado se puede realizar de varias maneras:

- Empleando instrucciones para obtener el tiempo (por ejemplo en Java System.currentTimeMillis() o System.nanoTime(), en C clock(), time(...)).
- Usando la orden Linux time.
- Usando herramientas para ello ("*profiler*") asociadas al lenguaje (por ejemplo, la opción -Xprof de java o gprof para C).
- Usando herramientas "*profiler*" asociadas al entorno de desarrollo.

Observaciones

En ocasiones la medición del tiempo no es muy precisa, sobre todo cuando el tamaño del vector es pequeño. Se recomienda tener esto en cuenta, para escoger las instrucciones que se utilizarán para medir tiempos y focalizar la comparación de los tiempos necesarios en el caso de vectores grandes, mientras que los vectores pequeños serán útiles para las comparaciones de operaciones elementales.

Se admiten valores repetidos en los vectores. En este caso la búsqueda parará cuando el elemento buscado se encuentre por primera vez.

Los valores que se buscarán se generarán de modo aleatorio, al igual que los elementos del vector. Se admite la posibilidad de que un elemento no aparezca en el vector, en cuyo caso también se tendrán en cuenta los datos de esta búsqueda para el análisis.

Los métodos de búsqueda deben implementarse y modificarse para realizar los cálculos necesarios para este análisis.

Normas de entrega

La práctica se realizará en **grupos de dos personas**. Se debe entregar el código fuente de los programas que se han creado para la realización de la práctica (se recomienda el uso del lenguaje Java aunque alternativamente se permite realizar en cualquier lenguaje de programación). Dichos programas deben compilar y ejecutar perfectamente en las máquinas de los laboratorios.

Estos programas deben generar los siguientes ficheros como mínimo para cada supuesto (búsqueda en vectores ordenados, búsqueda en vectores no ordenados):

- Fichero con los datos promedio: tamaño del vector, tiempo promedio de la búsqueda secuencial, tiempo promedio de la búsqueda binaria, operaciones promedio de la búsqueda secuencial, operaciones promedio de la búsqueda binaria
- Fichero con los datos individuales de cada iteración: tamaño del vector, tiempo de búsqueda secuencial, tiempo de búsqueda binaria, comparaciones secuencial, comparaciones binaria

Además se entregará un documento en **formato PDF** donde se indique lo siguiente:

- Nombre de ambos alumnos.
- Una breve descripción de cómo se ha realizado la práctica: rangos de tamaños y repetición de medidas, programas utilizados para el análisis, conclusiones obtenidas y razones que soportan estas conclusiones, etc.
- Las tablas y gráficos con los resultados obtenidos en las medidas experimentales de los algoritmos.
- Análisis teórico, indicando cómo se ha realizado, y la comparación que se obtiene en base a este análisis.
- La fórmula obtenida experimentalmente para cada uno de los algoritmos.
- **Una explicación razonada de qué algoritmo es mejor en cada caso, basada en los resultados experimentales.**
- **Una comparación razonada de los resultados obtenidos con el análisis teórico y los resultados experimentales.**

Fecha de entrega

La fecha límite de entrega es **el 7 de octubre a las 23:55**. Los ficheros de la práctica se empaquetarán en un fichero zip cuyo nombre será: p1_login1_login2.zip, donde login1 y login2 son los nombre de usuario en el laboratorio de los alumnos que forman el grupo. El fichero zip se entregará a través del Aula Virtual (<http://aulas.inf.uva.es>).