

Sorting Considerations Sorting 1

We consider sorting a list of records, either into ascending or descending order, based upon the value of some field of the record we will call the sort key.

The list may be contiguous and randomly accessible (e.g., an array), or it may be dispersed and only sequentially accessible (e.g., a linked list). The same logic applies in both cases, although implementation details will differ.

When analyzing the performance of various sorting algorithms we will generally consider two factors:

- the number of sort key comparisons that are required
- the number of times records in the list must be moved

Both worst-case and average-case performance is significant.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Internal or External? Sorting 2

In an internal sort, the list of records is small enough to be maintained entirely in physical memory for the duration of the sort.

In an external sort, the list of records will not fit entirely into physical memory at once. In that case, the records are kept in disk files and only a selection of them are resident in physical memory at any given time.

We will consider only internal sorting at this time.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Sortable Records Sorting 3

The records stored in the list may be simple (e.g., `string` or `int`) or they may be complex structures. In any case, we will assume that:

- there is an implicit conversion of a record into its key value,
- key values may be compared using the usual relational operators (`<`, `>`, etc.).

The first assumption requires that the record type implementation provides an appropriate conversion operator.

The second assumption requires that the key type implementation provides an overloading of the relational operators.

None of these features are difficult for a client of the sortable list to implement, and the assumptions will make the implementation of the sortable list considerably simpler and more elegant.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Sorted Lists Sorting 4

Sorting a list requires accessing the data elements stored in the list. For efficiency this suggests that the sort capability should be provided via member functions of the list class in question. We will follow that approach here, although the alternative is certainly feasible.

Building upon the `LinkedListT` or `ArrayT` classes discussed earlier, we could derive sorted variants, overriding some member functions of the base class and adding new member functions for sorting.

If we do that, what base member functions must be overridden?

- The insertion functions must now take into account the ordering of the list elements, placing each new element in the proper location.
- A number of the base member functions are unsuitable for a sorted list type and must be disabled.
- The search functions may now take advantage of the list element ordering.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Generic Linked List Interface

Sorting 5

```

template <class Item> class LinkListT {
protected:
    LinkNodeT<Item>* Head;    // points to head node in list
    LinkNodeT<Item>* Tail;   // points to tail node in list
    LinkNodeT<Item>* Curr;   // points to "current" node in list
public:
    LinkListT();
    LinkListT(const LinkListT<Item>& Source);
    LinkListT<Item>& operator=(const LinkListT<Item>& Source);
    ~LinkListT();
    bool isEmpty() const;
    bool inList() const;
    bool PrefixNode(Item newData);
    bool AppendNode(Item newData);
    bool InsertAfterCurr(Item newData);
    bool Advance();
    void gotoHead();
    void gotoTail();
    bool MakeEmpty();
    bool DeleteCurrentNode();
    bool DeleteValue(Item Target);
    Item getCurrentData() const;
    void PrintList(ostream& Out);
};

```

Unsuitable for
SortedList

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Ordered Insertion

Sorting 6

Inserting a new record into an ordered list requires:

- searching the list until the appropriate location is found
- creating an empty "slot" in the list to hold the new record
- placing the new record into that "slot"

The search phase is essentially trivial, aside from the issue of ties.

The creation of an empty "slot" depends upon the type of list:

- For a contiguous list, the tail of the list must be shifted to make room for the new element.
- For a linked list, a new list node must be allocated and inserted.

Placing the record into the "slot" is trivial.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Linked Sorted List Interface

Sorting 7

```

template <class Item> class SortedLinkListT {
private:
    LinkNodeT<Item>* Head;    // points to head node in list
    LinkNodeT<Item>* Tail;   // points to tail node in list
    LinkNodeT<Item>* Curr;   // points to "current" node in list
public:
    SortedLinkListT();
    SortedLinkListT(const SortedLinkListT<Item>& Source);
    SortedLinkListT<Item>& operator=(const SortedLinkListT<Item>& Source);
    ~SortedLinkListT();

    bool isEmpty() const;    // indicates if list is empty
    bool inList() const;    // indicates if list has a defined
                            // current location
    bool Find(Item toFind); // moves current location to first
                            // node containing matching data
    bool Insert(Item newData); // inserts new data element in proper
                            // location
    bool MakeEmpty();       // resets list to empty state
    bool DeleteCurrentNode(); // deletes node in current location
    bool DeleteValue(Item Target); // deletes first node containing
                            // a matching data element

    Item getCurrentData() const; // returns data element in current
                            // location
    void PrintList(ostream& Out); // prints formatted list to stream
};

```

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Efficiency Considerations

Sorting 8

The approach described so far maintains the list in sorted order at all times, by placing each new element in the appropriate location when it is inserted to the list.

For a list containing N elements, whether contiguous or linked, the average cost of ordered insertion of a single new element is $\Theta(N)$. When an exact analysis is done, performance is worse, of course, for a contiguous list.

Greater efficiency can be obtained if the list elements are inserted and then an efficient sort algorithm is applied to the list.

Which approach is preferred depends upon how often the list receives insertions, and whether those insertions tend to be grouped or isolated (in time).

We will now consider several general sorting algorithms that may be applied to a list of sortable records.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Sortable Linked List Interface Sorting 9

We derive a sortable linked class from the generic linked list class seen earlier:

```

// SortableLinkedList.h
//
// The SortableLinkedListT template class provides a simple implementation
// for a singly-linked list structure consisting of ListNodeT objects,
// and adds support for client-activated sorting.
//
// Assumptions:
// Item objects may be compared with the relational operators < and ==,
// and may be inserted to an output stream with <<.
//
#ifndef SORTABLELINKLISTT_H
#define SORTABLELINKLISTT_H
#include "LinkedList.h"

template <class Item> class SortableLinkedListT : public LinkedListT<Item> {
protected:
    // add private helper functions for sort support here
// ... continues ...
    
```

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Sortable Linked List Interface Sorting 10

```

// ... continued ...
public:
    SortableLinkedListT();
    SortableLinkedListT(const SortableLinkedListT<Item>& Source);
    SortableLinkedListT<Item>&
        operator=(const SortableLinkedListT<Item>& Source);
    ~SortableLinkedListT();

    // add public interface functions for sorting here
};

#include "SortableLinkedListT.cpp"
#endif
    
```

The idea is to provide a flexible, somewhat traditional base linked list class to which we may add any sort functions we like.

The remainder of this chapter will be devoted primarily to considering the various sort algorithms we could plug into this interface.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Insertion Sort Sorting 11

Insertion Sort:

Insertion sort closely resembles the insertion function for a sorted list.

For a contiguous list, the primary costs are the comparisons to determine which part of the sorted portion must be shifted, and the assignments needed to accomplish that shifting of the sorted tail.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Insertion Sort Average Comparisons Sorting 12

Assuming a list of N elements, Insertion Sort requires:

Average case: $N^2/4 + O(N)$ comparisons and $N^2/4 + O(N)$ assignments

Consider the element which is initially at the K^{th} position and suppose it winds up at position j , where j can be anything from 1 to K . A final position of j will require $K - j + 1$ comparisons. Therefore, on average, the number of comparisons to place the K^{th} element is:

$$\frac{1}{K} \sum_{j=1}^K (K - j + 1) = \frac{1}{K} \left[K^2 - \frac{K(K+1)}{2} + K \right] = \frac{K+1}{2}$$

The average total cost for insertion sort on a list of N elements is thus:

$$\sum_{k=2}^N \left(\frac{K+1}{2} \right) = \sum_{k=1}^{N-1} \left(\frac{K+2}{2} \right) = \frac{1}{2} \left[\frac{(N+1)(N+2)}{2} \right] = \frac{1}{4} N^2 + \frac{3}{4} N - 1$$

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Insertion Sort Average Assignments Sorting 13

(...continued...)

The analysis for assignments is similar, differing only in that if the element in the K^{th} position winds up at position j , where j is between 1 to $K - 1$ inclusive, then the number of assignments is $K - j + 2$. The case where the element does not move is special, in that no assignments take place.

Proceeding as before, the average total number of assignments satisfies:

$$\sum_{k=2}^N \left(\frac{K+3}{2} - \frac{2}{K} \right) < \sum_{k=1}^{N-1} \left(\frac{K+4}{2} - \frac{1}{4} \right) = \frac{1}{4}N^2 + \frac{7}{4}N + 3$$

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Insertion Sort Worst/Best Case Analysis Sorting 14

Worst case: $N^2/2 + O(N)$ comparisons and $N^2/2 + O(N)$ assignments

QTP: when will the worst case be achieved?

Best case: $N - 1$ comparisons and no assignments (list is pre-sorted)

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Lower Bound on the Cost of Sorting Sorting 15

Before considering how to improve on Insertion Sort, consider the question:

How fast is it possible to sort?

Now, “fast” here must refer to algorithmic complexity, not time. We will consider the number of comparisons of elements a sorting algorithm must make in order to fully sort a list.

Note that this is an extremely broad issue since we seek an answer of the form: any sorting algorithm, no matter how it works, must perform at least $O(f(N))$ comparisons when sorting a list of N elements.

Thus, we cannot simply consider any particular sorting algorithm...

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Possible Orderings of N Elements Sorting 16

A bit of combinatorics (the mathematics of counting)...

Given a collection of N distinct objects, the number of different ways to line them up in a row is $N!$.

Thus, a sorting algorithm must, in the worst case, determine the correct ordering among $N!$ possible results.

If the algorithm compares two elements, the result of the comparison eliminates certain orderings as the final answer, and directs the “search” to the remaining, possible orderings.

We may represent the process of comparison sorting with a binary tree...

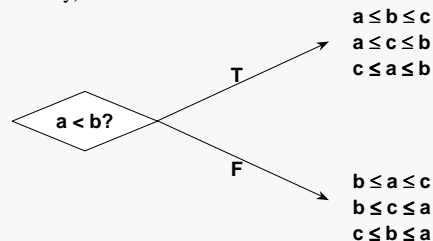
Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Comparison Trees

Sorting 17

A comparison tree is a binary tree in which each internal node represents the comparison of two particular elements of a set, and the edges represent the two possible outcomes of that comparison.

For example, given the set $A = \{a, b, c\}$ a sorting algorithm will begin by comparing two of the elements (it doesn't matter which two, so we'll choose arbitrarily):



Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Lower Bound on Comparisons

Sorting 18

Theorem: Any algorithm that sorts a list of N entries by use of key comparisons must, in its worst case, perform at least $\log(N!)$ comparisons of keys, and, in the average case, it must perform at least $\log(N!)$ comparisons of keys.

proof: The operation of the algorithm can be represented by a comparison tree. That tree must have $N!$ leaves, since there are $N!$ possible answers.

We proved earlier that the number of levels in a binary tree with $N!$ leaves must have at least $1 + \lceil \log(N!) \rceil$. If the answer corresponds to a leaf in the bottom level of the comparison tree, the algorithm must traverse at least $\lceil \log(N!) \rceil$ internal nodes, and so must do at least $\lceil \log(N!) \rceil$ comparisons.

That proves the worst case lower bound.

The proof of the average case lower bound is tediously similar.

QED

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Stirling's Formula

Sorting 19

The average case lower bound, $\log(N!)$ comparisons, is somewhat ugly. We can simplify it by applying a result known as Stirling's Formula:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left[1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right]$$

or, converting to logarithmic form:

$$\ln(n!) \approx \left(n + \frac{1}{2}\right) \ln n - n + \frac{1}{2} \ln(2\pi) + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)$$

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Simplified Lower Bound

Sorting 20

Using Stirling's Formula, and changing bases, we have that:

$$\begin{aligned} \log(n!) &\approx \left(n + \frac{1}{2}\right) \log n - \log(e)n + \frac{1}{2} \log(2\pi) + \frac{\log(e)}{12n} \\ &\approx \left(n + \frac{1}{2}\right) \left(\log n - \frac{3}{2}\right) + 2 \\ &= n \log n - \frac{3}{2}n + \frac{1}{2} \log n + \frac{5}{4} \end{aligned}$$

For most practical purposes, only the first term matters here, so you will often see the assertion that the lower bound for comparisons in sorting is $n \log n$.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Improving Insertion Sort Sorting 21

Insertion sort is most efficient when the initial position of each list element is fairly close to its final position (take another look at the analysis).

Consider:

10	8	6	20	4	3	22	1	0	15	16
----	---	---	----	---	---	----	---	---	----	----

Pick a step size (5 here) and logically break the list into parts.

10					3					16
	8					22				
		6					1			
			20					0		
				4						15

Sort the elements in each part. Insertion Sort is acceptable since it's efficient on short lists.

3					10					16
	8					22				
		1					6			
			0					20		
				4						15

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Improving Insertion Sort Sorting 22

This gives us:

3	8	1	0	4	10	22	6	20	15	16
---	---	---	---	---	----	----	---	----	----	----

Now we pick a smaller increment, 3 here, and repeat the process:

Partition list:

3				0				22			15	
	8				4				6			16
		1				10				20		

Sort the parts:

0				3				15				22
	4				6				8			16
		1				10				20		

Giving:

0	4	1	3	6	10	15	8	20	22	16
---	---	---	---	---	----	----	---	----	----	----

Finally repeat the process with step size 1:

0	1	3	4	6	8	10	15	16	20	22
---	---	---	---	---	---	----	----	----	----	----

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Why Is This Better? Sorting 23

Well...

- Until the last pass, the sublists that are being sorted are much shorter than the entire list — sorting two lists of length 1000 is faster than sorting one list of length 2000 .
- Since the sublists exhibit mixing as we change the step size, the effect of the early passes is to move each element closer to its final position in the fully sorted list.
- In the last pass, most of the elements are probably not too far from their final positions, and that's one situation where Insertion Sort is quite efficient.

QTP: Suppose that a sorting algorithm is, on average, $\Theta(N^2)$. Using that fact, how would the expected time to sort a list of length 50 compare to the time required to sort a list of length 100?

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Shell Sort Sorting 24

The process just shown is known as Shell Sort (after its originator: Donald Shell, 1959), and may be summed up as follows:

- partition the list into discontinuous sublists whose elements are some step size, h , apart.
- sort each sublist by applying Insertion Sort (or ...)
- decrease the value of h and repeat the first two steps, stopping after a pass in which h is 1.

The end result will be a sorted list because the final pass (with h being 1) is simply an application of Insertion Sort to the entire list.

What is an optimal sequence of values for the step size h ? No one knows...

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Shell Sort Performance Sorting 25

What is the performance analysis for Shell Sort? No one knows... in general.

Knuth has shown that if two particular step sizes are used, then Shell Sort takes $O(N^{5/2})$ time (versus $O(N^2)$ for simple Insertion Sort).

$\left(\frac{16N}{\pi}\right)^{1/3}$ and 1

Good mixing of the sublists can be provided by choosing the step sizes by the rule:

$$\begin{cases} h_1 = 1 \\ h_{i+1} = 3h_i + 1 \text{ for } i \geq 1 \end{cases}$$

Empirically, for large values of N, Shell Sort appears to be about $O(N^{5/4})$ if the step size scheme given above is used.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Divide and Conquer Sorting Sorting 26

Shell Sort represents a "divide-and-conquer" approach to the problem.

That is, we break a large problem into smaller parts (which are presumably more manageable), handle each part, and then somehow recombine the separate results to achieve a final solution.

In Shell Sort, the recombination is achieved by decreasing the step size to 1, and physically keeping the sublists within the original list structure. Note that Shell Sort is better suited to a contiguous list than a linked list.

We will now consider a somewhat similar algorithm that retains the divide and conquer strategy, but which is better suited to linked lists.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Merge Sort Sorting 27

In Merge Sort, we chop the list into two (or more) sublists which are as nearly equal in size as we can achieve, sort each sublist separately, and then carefully merge the resulting sorted sublists to achieve a final sorting of the original list.

Merging two sorted lists into a single sorted list is relatively trivial:

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Merge Sort Implementation Sorting 28

The public interface function is just a shell to call a recursive helper function:

```

template <class Item> void SortableLinkedListT<Item>::MergeSort() {
    MergeSortHelper(Head);
    Sorted = true;
}

template <class Item> void
SortableLinkedListT<Item>::MergeSortHelper(LinkNodeT<Item>* sHead) {
    if ( (sHead != NULL) && (sHead->getNext() != NULL) ) {
        LinkNodeT<Item>* SecondHalf = DivideFrom(sHead);
        MergeSortHelper(sHead);
        MergeSortHelper(SecondHalf);
        sHead = Merge(sHead, SecondHalf);
    }
}
    
```

partition list

sort each half

merge the pieces

In turn, the helper function uses two other private helper functions...

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Merge Sort Partition

Sorting 29

The partition divides the list nodes as evenly as possible:

```

template <class Item> LinkNodeT<Item>*
SortableView::DivideFrom(LinkNodeT<Item>* sHead) {

    LinkNodeT<Item> *Position,
                    *MidPoint = sHead,
                    *SecondHalf;

    if (MidPoint == NULL) return NULL;

    Position = MidPoint->getNext();
    while (Position != NULL) {
        Position = Position->getNext();
        if (Position != NULL) {
            MidPoint = MidPoint->getNext();
            Position = Position->getNext();
        }
    }
    SecondHalf = MidPoint->getNext();

    MidPoint->setNext(NULL);

    return SecondHalf;
}
    
```

Sublist is empty, so quit...

Walk Position to the end of the list, moving MidPoint at half the speed of Partition, so MidPoint winds up at the middle of the list.

Get head of second half of list.

Break sublist at it's middle.

Use a dummy head node for the merged list.

On each pass, merge the sublist head node which contains the smaller value.

Pick up nonempty sublist tail, if any.

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

Merge Sort Merge

Sorting 30

The public interface function is just a shell to call a recursive helper function:

```

template <class Item> LinkNodeT<Item>*
SortableView::Merge(LinkNodeT<Item>* First,
                    LinkNodeT<Item>* Second) {

    LinkNodeT<Item> *LastSorted;
    LinkNodeT<Item> Combined;
    LastSorted = &Combined;

    while ( (First != NULL) && (Second != NULL) ) {
        if (First->getData() <= Second->getData()) {
            LastSorted->setNext(First);
            LastSorted = First;
            First = First->getNext();
        }
        else {
            LastSorted->setNext(Second);
            LastSorted = Second;
            Second = Second->getNext();
        }
    }
    if (First == NULL) LastSorted->setNext(Second);
    else LastSorted->setNext(First);
    return ( Combined.getNext() );
}
    
```

Sublist is empty, so quit...

Walk Position to the end of the list, moving MidPoint at half the speed of Partition, so MidPoint winds up at the middle of the list.

Get head of second half of list.

Break sublist at it's middle.

Use a dummy head node for the merged list.

On each pass, merge the sublist head node which contains the smaller value.

Pick up nonempty sublist tail, if any.

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

Merge Sort Performance

Sorting 31

At a high level, the implementation involves two activities, partitioning and merging, each represented by a corresponding function.

The number of partitioning steps equals the number of merge steps, partitioning taking place during the recursive descent and merging during the recursive ascent as the calls back out.

Consider partitioning for a list of 8 elements:

14	32	8	23	4	9	19	10
14 32 8 23				4 9 19 10			
14 32		8 23		4 9		19 10	
14	32	8	23	4	9	19	10

The maximum number of levels is $\lceil \log N \rceil$.

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

Merge Sort Performance

Sorting 32

All the element comparisons take place during the merge phase. Logically, we may consider this as if the algorithm re-merged each level before proceeding to the next:

14	32	8	23	4	9	19	10
14 32		8 23		4 9		10 19	
8 14 23 32				4 9 10 19			
4 8 9 10 14 19 23 32							

So merging the sublists involves $\lceil \log N \rceil$ passes. On each pass, each list element is used in (at most) one comparison, so the number of element comparisons per pass is N . Hence, the number of comparisons for Merge Sort is $O(N \lceil \log N \rceil)$.

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

Merge Sort Summary Sorting 33

Merge Sort comes very close to the theoretical optimum number of comparisons. A closer analysis shows that for a list of N elements, the average number of element comparisons using Merge Sort is actually:

$$O(N \log N - 1.1583N + 1)$$

Recall that the theoretical minimum is: $N \log N - 1.44N + O(N)$

For a linked list, Merge Sort is the sorting algorithm of choice, providing nearly optimal comparisons and requiring NO element assignments (although there is a considerable amount of pointer manipulation), and requiring NO significant additional storage.

For a contiguous list, Merge Sort would require either using $O(N)$ additional storage for the sublists or using a considerably complex algorithm to achieve the merge with a small amount of additional storage.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Heap Sort Sorting 34

A list can be sorted by first building it into a heap, and then iteratively deleting the root node from the heap until the heap is empty. If the deleted roots are stored in reverse order in an array they will be sorted in ascending order (if a max heap is used).

```

void HeapSort(int* List, int Size) {
    HeapT<int> toSort(List, Size);
    toSort.BuildHeap();

    int Idx = Size - 1;
    while ( !toSort.isEmpty() ) {
        List[Idx] = toSort.RemoveRoot();
        Idx--;
    }
}
    
```

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Cost of Deleting the Roots Sorting 35

Recalling the earlier analysis of building a heap, level k of a full and complete binary tree will contain 2^k nodes, and that those nodes are k levels below the root level.

So, when the root is deleted the maximum number of levels the swapped node can sift down is the number of the level from which that node was swapped.

Thus, in the worst case, for deleting all the roots...

$$\begin{aligned} \text{Comparisons} &= \sum_{k=1}^{d-1} k 2^k = 4 \sum_{k=1}^{d-1} k 2^{k-1} = 4 \left[(d-2)2^{d-1} + 1 \right] \\ &= 2N \log N + 2 \lceil \log N \rceil - 4N \end{aligned}$$

As usual, with Heap Sort, this would entail half as many element swaps.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Heap Sort Performance Sorting 36

Adding in the cost of building the heap from our earlier analysis,

$$\begin{aligned} \text{Total Comparisons} &= (2N - 2 \log N) + (2N \lceil \log N \rceil + 2 \lceil \log N \rceil - 4N) \\ &= 2N \lceil \log N \rceil - 2N \end{aligned}$$

and...

$$\text{Total Swaps} = N \log N - N$$

So, in the worst case, Heap Sort is $\Theta(N \log N)$ in both swaps and comparisons.

The average case analysis is apparently so difficult no one has managed it rigorously. However, empirical analysis indicates that Heap Sort does not do much better, on average, than in the worst case.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

QuickSort

Sorting 37

QuickSort is conceptually similar to MergeSort in that it is a divide-and-conquer approach, involving successive partitioning of the list to be sorted.

The differences lie in the manner in which the partitioning is done, and that the sublists are maintained in proper relative order so no merging is required.

QuickSort is a naturally recursive algorithm, each step of which involves:

- pick a pivot value for the current sublist
- partition the sublist into two sublists, the left containing only values less than the pivot value and the right containing only values greater than or equal to the pivot value
- recursively process each of the resulting sublists

As they say, the devil is in the details...

Computer Science Dept. Va Tech. July 2000 Data Structures & File Management ©2000 McQuain WD

Importance of the Pivot Value

Sorting 38

The choice of the pivot value is crucial to the performance of QuickSort. Ideally, the partitioning step produces two equal sublists, as here:

43	71	87	14	53	38	90	51	41
----	----	----	----	----	----	----	----	----

↙

In the worst case, the partitioning step produces one empty sublist, as here:

43	71	87	14	53	38	90	51	41
----	----	----	----	----	----	----	----	----

↘

Theoretically, the ideal pivot value is the median of the values in the sublist; unfortunately, finding the median is too expensive to be practical here.

Computer Science Dept. Va Tech. July 2000 Data Structures & File Management ©2000 McQuain WD

Choosing the Pivot Value

Sorting 39

Commonly used alternatives to finding the median are:

- take the value at some fixed position (first, middle-most, last, etc.)
- take the median value among the first, last and middle-most values
- find three distinct values and take the median of those

The third does not guarantee good performance, but is the best of the listed strategies since it is the only one that guarantees two nonempty sublists. (Of course, if you can't find three distinct values, this doesn't work, but in that case the current sublist doesn't require any fancy sorting — a quick swap will finish it off efficiently.)

QTP: under what conditions would choosing the first value produce consistently terrible partitions?

Each of the given strategies for finding the pivot is $O(1)$ in comparisons.

Computer Science Dept. Va Tech. July 2000 Data Structures & File Management ©2000 McQuain WD

Partitioning a Sublist Efficiently

Sorting 40

Since each iteration of QuickSort requires partitioning a sublist, this must be done efficiently. Fortunately, there is a simple algorithm for partitioning a sublist of N elements that is $O(N)$ in comparisons and assignments:

```

template <class Item> int SortableArrayT<Item>::Partition(int Lo, int Hi) {
    Item Pivot;
    int Idx,
        LastPreceder;
    Swap(List[Lo], List[(Lo + Hi)/2]); // take middle-most element as Pivot
    Pivot = List[Lo]; // move it to the front of the list
    LastPreceder = Lo;

    for (Idx = Lo + 1; Idx <= Hi; Idx++) {
        if (List[Idx] < Pivot) {
            LastPreceder++;
            Swap(List[LastPreceder], List[Idx]);
        }
    }
    Swap(List[Lo], List[LastPreceder]);

    return LastPreceder;
}

```

Computer Science Dept. Va Tech. July 2000 Data Structures & File Management ©2000 McQuain WD

QuickSort Implementation

Sorting 41

Assuming the pivot and partition function just described, the main QuickSort function is quite trivial:

```
template <class Item> void SortableArrayT<Item>::QuickSort() {
    QuickSortHelper(0, Usage - 1);
}

template <class Item>
void SortableArrayT<Item>::QuickSortHelper(int Lo, int Hi) {
    int PivotIndex;
    if (Lo < Hi) {
        PivotIndex = Partition(Lo, Hi);
        QuickSortHelper(Lo, PivotIndex - 1); // recurse on lower part, then
        QuickSortHelper(PivotIndex + 1, Hi); // on higher part
    }
}
```

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

QuickSort Performance

Sorting 42

As with Merge Sort, we can view QuickSort as consisting of a sequence of steps, each involving pivot selection and then partitioning of each nontrivial sublist.

What is the total partitioning cost at each step?

As implemented here, each partitioning step isolates a copy of the pivot value and that value is not subsequently moved, so the number of elements that are subject to partitioning decreases at each step. However, it is relatively clear that the total cost of partitioning at any step is no worse than $O(N)$.

But, how many steps must be performed?

Ah, there's the rub... that depends upon how well the pivot values work.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

QuickSort Performance Analysis

Sorting 43

For simplicity in analysis, we will make the following assumptions:

- the key values to be partitioned are $1 \dots N$
- $C(N)$ is the number of comparisons made by QuickSort when applied to a list of length N
- $S(N)$ is the number of swaps made by QuickSort when applied to a list of length N

If partitioning produces one sublist of length r and one of $N - r - 1$, then:

$$C(N) = N - 1 + C(r) + C(N - r - 1)$$

since the initial partition will require $N - 1$ element comparisons.

The formula derived here is known as a recurrence relation.

For a given value of r , it is possible to solve for $C(N)$...

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

QuickSort Performance Worst Case

Sorting 44

Suppose $r = 0$; i. e., that QuickSort produces one empty sublist and merely splits off the pivot value from the remaining $N - 1$ elements. Then we have:

$$\begin{aligned} C(1) &= 0 \\ C(2) &= 1 + C(1) &&= 1 \\ C(3) &= 2 + C(2) &&= 2 + 1 \\ C(4) &= 3 + C(3) &&= 3 + 2 + 1 \\ &\dots \\ C(N) &= N - 1 + C(N - 1) &&= (N - 1) + (N - 2) + \dots + 1 \\ &&&= 0.5N^2 - 0.5N \end{aligned}$$

This is the worst case, and is as bad as the worst case for selection sort.

Similar analysis shows that the number of swaps in this case is:

$$S(N) = 0.5N^2 + 1.5N - 1$$

which is $O(1.5N^2)$ element assignments, even worse than insertion sort.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

QuickSort Performance Average Case Sorting 45

For the average case, we will consider all possible results of the partitioning phase and compute the average of those. We assume that all possible orderings of the list elements are equally likely to be the correct sorted order, and let p denote the pivot value chosen.

Thus, after partitioning, recalling we assume the key values are $1 \dots N$, the values $1, \dots, p-1$ are to the left of p , and the values $p+1, \dots, N$ are to the right of p .

Let $S(N)$ be the average number of swaps made by QuickSort on a list of length N , and $S(N, p)$ be the average number of swaps if the value p is chosen as the initial pivot.

The partition implementation given here will perform $p-1$ swaps within the loop, and two more outside the loop. Therefore...

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

QuickSort Performance Average Case Sorting 46

$$S(N, p) = p + 1 + S(p-1) + S(N-p)$$

Now there are N possible choices for p ($1 \dots N$), and so if we sum this expression from $p = 1$ to $p = N$ and divide by N we have:

$$S(N) = \frac{N}{2} + \frac{3}{2} + \frac{2}{N} [S(0) + S(1) + \dots + S(N-1)] \quad (A)$$

That's another recurrence relation... we may solve this by playing a clever trick; first note that if QuickSort were applied to a list of $N-1$ elements, we'd have:

$$S(N-1) = \frac{N-1}{2} + \frac{3}{2} + \frac{2}{N-1} [S(0) + S(1) + \dots + S(N-2)] \quad (B)$$

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

QuickSort Performance Average Case Sorting 47

If we multiply both sides of equation (A) by N , and multiply both sides of equation (B) by $N-1$, and then subtract, we can obtain:

$$NS(N) - (N-1)S(N-1) = N + 1 + 2S(N-1) \quad (C)$$

Rearranging terms, we get:

$$\frac{S(N)}{N+1} = \frac{S(N-1)}{N} + \frac{1}{N} \quad (D)$$

A closed-form solution for this recurrence relation can be guessed by writing down the first few terms of the sequence. Doing so, we obtain...

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

QuickSort Performance Average Case Sorting 48

$$\frac{S(N)}{N+1} = \frac{S(2)}{3} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{N} \quad (E)$$

Now, it can be shown that for all $N \geq 1$,

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{N} = \ln N + O(1)$$

Applying that fact to equation (E), we can show that: $\frac{S(N)}{N+1} = \ln N + O(1)$

Algebra and a log base conversion yield:

$$S(N) = N \ln N + O(N) \approx 0.69N \log N + O(N)$$

Computer Science Dept Va Tech July 2000
Data Structures & File Management
©2000 McQuain WD

QuickSort Performance Average Case

Sorting 49

A similar analysis for comparisons starts with the relationship:

$$C(N, p) = N - 1 + C(p - 1) + C(N - p)$$

This eventually yields:

$$C(N) = 2N \ln N + O(N) \approx 1.39N \log N + O(N)$$

So, on average, the given implementation of QuickSort is $O(N \log N)$ in both swaps and comparisons.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

QuickSort Improvements

Sorting 50

QuickSort can be improved by

- switching to a nonrecursive sort algorithm on sublists that are relatively short. The common rule of thumb is to switch to insertion sort if the sublist contains fewer than 10 elements.
- eliminating the recursion altogether. However, the resulting implementation is considerably more difficult to understand (and therefore to verify) than the recursive version.
- making a more careful choice of the pivot value than was done in the given implementation.
- improving the partition algorithm to reduce the number of swaps. It is possible to reduce the number of swaps during partitioning to about 1/3 of those used by the given implementation.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Comparison of Techniques

Sorting 51

The earlier sorting algorithms are clearly only suitable for short lists, where the total time will be negligible anyway.

While MergeSort can be adapted for contiguous lists, and QuickSort for linked lists, each adaptation induces considerable loss of efficiency and/or clarity; therefore, MergeSort and QuickSort cannot really be considered competitors.

HeapSort is inferior to the average case of QuickSort, but not by much. In the worst cases, QuickSort is among the worst sorting algorithms discussed here. From that perspective, HeapSort is attractive as a hedge against the possible worst behavior of QuickSort.

However, in practice QuickSort rarely degenerates to its worst case and is almost always $O(N \log N)$.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Beating the $N \log(N)$ Barrier

Sorting 52

Theoretical analysis tells us that no sorting algorithm can require fewer comparisons than $O(N \log N)$, right?

Well, actually no... the derivation of that result assumes that the algorithm works by comparing key values.

It is possible to devise sorting algorithms that do not compare the elements of the list to each other. In that case, the earlier result does not apply.

However, such algorithms require special assumptions regarding the type and/or range of the key values to be sorted.

Computer Science Dept Va Tech July 2000

Data Structures & File Management

©2000 McQuain WD

Bin Sort Sorting 53

Assume we need to sort a list of integers in the range 0-99:

34	16	83	76	40	72	38	80	89	87
----	----	----	----	----	----	----	----	----	----

Given an integer array of dimension 100 (the bins) for storage, make a pass through the list of integers, and place each into the bin that matches its value:

$$\text{Bin}[\text{Source}[k]] = \text{Source}[k]$$

This will take one pass, requiring $O(N)$ work, much better than $N \log(N)$.

Limitations of Bin Sort:

- the number of bins is determined by the range of the values
- only suitable for integer-like values

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Radix Sort Sorting 54

Assume we need to sort a list of integers in the range 0-99:

34	16	83	76	40	72	38	80	89	87
----	----	----	----	----	----	----	----	----	----

Given an array of 10 linked lists (bins) for storage, make a pass through the list of integers, and place each into the bin that matches its 1's digit.

Then, make a second pass, taking each bin in order, and append each integer to the bin that matches its 2's digit.

Bin									
0:	40	80							
1:									
2:	72								
3:	83								
4:	34								
5:									
6:	16	76							
7:	87								
8:	38								
9:	89								

Bin									
0:									
1:	16								
2:									
3:	34	38							
4:	40								
5:									
6:									
7:	72	76							
8:	80	83	87	89					
9:									

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD

Radix Sort Analysis Sorting 55

Now if you just read the bins, in order, the elements will appear in ascending order. Each pass takes $O(N)$ work, and the number of passes is just the number of digits in the largest integer in the original list.

That beats $N \log(N)$, by a considerable margin, but only in a somewhat special case.

Bin									
0:									
1:	16								
2:									
3:	34	38							
4:	40								
5:									
6:									
7:	72	76							
8:	80	83	87	89					
9:									

Note the number of bins is determined by the number of possible values in each position, while the number of passes is determined by the maximum length of the data elements.

Computer Science Dept Va Tech July 2000 Data Structures & File Management ©2000 McQuain WD