Practical Session No. 11

Linear Time Sort Algorithms

**Counting-Sort**
A sort algorithm not based on comparisons and supports duplicate keys.
- A is an input array of length n
- B is the output array.
- C is an auxiliary array of size k.
- **Assumption:** A consists of elements with integer keys in the range [1..k]

**Counting-Sort (A, B, k)**

for i ← 1 to k
  C[i] ← = 0
for j ← 1 to n
  C[A[j]] ← C[A[j]] + 1
  // C[i] = the number of appearances of i in A.
for i ← 2 to k
  C[i] ← C[i] + C[i-1]
  // C[i] = the number of elements in A that are ≤ i
for j ← n downto 1
  B[C[A[j]]] ← A[j]
  C[A[j]] ← C[A[j]] - 1
return B

**Example:**
n=6, k=3

<table>
<thead>
<tr>
<th>Index</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

For each iteration:
  - B = 1
  - C = 1 3 6
  - B = 1 1 1 3 3 3
  - C = 0 2 3
  - B = 1 1 1 2 3 3
  - C = 0 3 4
  - B = 1 1 1 2 3 3
  - C = 0 2 4
  - B = 1 1 2 3 3 3
  - C = 0 2 3
  - B = 1 1 2 3 3 3
  - C = 0 2 3
**Run-time Complexity**: \(O(n+k)\)

This is an improvement on comparison-based sorts, which need \(n*\log n\) time. Counting sort is **stable**, two elements with the same key value will appear in the output in the same order as they appeared in the input. Stability is important when there is additional data besides the key.

**Radix-Sort**

A stable sort algorithm for sorting elements with \(d\) digits, where digits are in base \(b\), i.e., in range \([0,b-1]\).

The algorithm uses a stable sort algorithm to sort the keys by each digit, starting with the least significant digit (the rightmost).

**Radix-Sort**

\[
\text{Radix-Sort}(A[1..n], d) \quad \text{מיון בסיס}
\]

\[
\text{for } i \leftarrow 1 \text{ to } d
\]
\[
\quad \text{Use stable sort to sort } A \text{ on digit } i \text{ (like counting-sort)}
\]

**Run-time Complexity**

Assuming the stable sort runs in \(O(n+b)\) (such as counting sort) the running time is \(O(d(n+b)) = O(dn+db)\).

If \(d\) is constant and \(b=O(n)\), the running time is \(O(n)\).

**Example**

7 numbers with 3 digits. \(n=7\), \(d=3\), \(b=10\) (decimal)

\[329, 457, 657, 839, 436, 720, 355\]

\[
720, 355, 436, 457, 657, 329, 839 \quad \text{– sorted by digit 1}
\]

\[
720, 329, 436, 839, 355, 457, 657 \quad \text{– sorted by digit 2 (and 1)}
\]

\[
329, 355, 436, 457, 657, 720, 839 \quad \text{– sorted}
\]

**Bucket-Sort**

- Assumption: Input array elements are uniformly distributed over the interval \([0,1]\).

The idea of bucket-sort is to divide the interval \([0,1)\) into \(n\) equal-sized subintervals (buckets), and then distribute the \(n\) input numbers into the buckets. To produce the output we simply sort the elements in each bucket and then creating a sorted list by going through the buckets in order.

**Bucket-sort (A)**

\[
\text{n } \leftarrow \text{ length}(A)
\]

\[
\text{for } i \leftarrow 1 \text{ to } n \text{ do}
\]
\[
B\left[\lfloor nA[i] \rfloor \right] \leftarrow A[i]
\]

\[
\text{for } i \leftarrow 1 \text{ to } n \text{ do}
\]
\[
\quad \text{sort } B[i] \text{ with insertion sort}
\]
\[
\quad \text{concatenate the lists } B[0], B[1], \ldots, B[n-1] \text{ together in order}
\]

**Run-time Complexity**

Assuming the inputs are uniformly distributed over \([0,1]\), we expect \(O(1)\) elements in each bucket (average case), thus sorting them takes \(O(1)\) expected time.

We insert \(n\) elements into \(n\) buckets in \(O(n)\) and we concatenate the Lists in \(O(n)\) ⇒ Total expected run time: \(O(n)\).
Sort Algorithms Review

<table>
<thead>
<tr>
<th>Keys Type</th>
<th>Average run-time</th>
<th>Worst case run-time</th>
<th>Extra Space</th>
<th>In Place</th>
<th>Stable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion Sort</td>
<td>any</td>
<td>O(n^2)</td>
<td>O(n^2)</td>
<td>O(1)</td>
<td>√</td>
</tr>
<tr>
<td>Merge Sort</td>
<td>any</td>
<td>O(nlogn)</td>
<td>O(nlogn)</td>
<td>O(n)</td>
<td>X</td>
</tr>
<tr>
<td>Heap Sort</td>
<td>any</td>
<td>O(nlogn)</td>
<td>O(nlogn)</td>
<td>O(1)</td>
<td>√</td>
</tr>
<tr>
<td>Quick Sort</td>
<td>any</td>
<td>O(nlogn)</td>
<td>O(n^2)</td>
<td>O(1)</td>
<td>√</td>
</tr>
<tr>
<td>Counting Sort</td>
<td>integers [1..k]</td>
<td>O(n+k)</td>
<td>O(n+k)</td>
<td>O(n+k)</td>
<td>X</td>
</tr>
<tr>
<td>Radix Sort</td>
<td>d digits in base b</td>
<td>O(d(b+n))</td>
<td>O(d(b+n))</td>
<td>Depends on the stable sort used</td>
<td>Depends on the stable sort used</td>
</tr>
<tr>
<td>Bucket sort</td>
<td>[0,1)</td>
<td>O(n)</td>
<td>O(n^2)</td>
<td>O(n)</td>
<td>X</td>
</tr>
</tbody>
</table>

**Question 1**

There are \( n \) integers in range \([1...k]\). Suggest an algorithm that preprocesses the input in \( O(n+k) \) time and then returns how many numbers there are in range \([a...b]\) in \( O(1) \) time for any given \( a \) and \( b \).

**Solution**

The algorithm is based on counting-sort:

```plaintext
Preprocess(A, k)
    for i ← 1 to k
        C[i] ← 0
    for j ← 1 to n
        C[A[j]] ← C[A[j]] + 1 // C[i] = the number of appearances of i in A.
    for i ← 2 to k
        C[i] ← C[i] + C[i-1] // C[i] = the number of elements in A that are ≤ i
    return(C)

Range(A,k,a,b)
    C= Preprocess(A,k);
    return(C[b]-C[a-1])
```

```plaintext
```
**Question 2**

Design an algorithm for sorting $n$ data items with keys in the range $[x, x+d)$ that runs in expected $O(n)$ time if the items are uniformly distributed over $[x, x+d]$, and runs in $O(n \log n)$ in the worst distribution case.

**Solution 1:**

Use bucket sort over the range $[x, x+d)$ with the following changes:

1. The elements in each bucket are stored in a AVL tree (instead of a linked list)
2. In the last stage, concatenate all the *inorder* visits of all the buckets one after another.

Note: the bucket distribution function will be $B \left\lfloor n((A[i] - x)/d) \right\rfloor \leftarrow A[i]$.

**Time Complexity:**

Let $n_i$ be the number of elements in the tree in bucket $i$.

1. Inserting the $n$ elements into the buckets takes $O(n_1 \log n_1 + n_2 \log n_2 + \ldots + n_n \log n_n)$
   - When the keys are uniformly distributed $n_i = O(1)$ for every $i$, hence
     $O(n_1 \log n_1 + n_2 \log n_2 + \ldots + n_n \log n_n) \leq c(n_1 + n_2 + \ldots + n_n) = cn$, where $c$ is a constant.
   - In the worst distribution cases:
     $O(n_1 \log n_1 + n_2 \log n_2 + \ldots + n_n \log n_n) \leq O(n_1 \log + n_2 \log + \ldots + n_n \log n_n) = O(n \log n)$
2. Inorder traversals of all buckets takes $O(n_1 + n_2 + \ldots + n_n) = O(n)$
3. Concatenation of all inorder traversal lists takes $O(n)$

The algorithm runs in $O(n)$ time for uniformly distributed keys and runs in $O(n \log n)$ in the worst distribution case.

**Solution 2:**

Execute in parallel the following two algorithms:

1. Original bucket sort
2. Any sort algorithm that takes $O(n \log n)$

Stop when one of the algorithms has stopped and return the sorted elements.

**Question 3**

Given a set of $n$ integers in the range $[1, n^3]$, suggest an efficient sorting algorithm.

**Solution:**

- Comparison-based algorithm takes $O(n \log n)$.
- Counting-sort: $k = n^3 \Rightarrow O(n+ n^3) = O(n^3)$
- Radix-sort: $b=n$, $d=4$, $d(b+n)=O(4(n+n))=O(n)$

Why is that?

Use radix-sort after preprocessing:

1. Convert all numbers to base $n$ in $O(n)$ total time using mod and div operations.
   
   $x = [x_3, x_2, x_1, x_0]$ ($x_0 = x \mod n$, $x_i = \lfloor x/n^i \rfloor \mod n$)
2. Call radix-sort on the transformed numbers. $O(n)$

All the numbers are in range 1 to $n^2$, therefore, we have at most 4 digits for each number. The running time for the suggest algorithm: $d=4$, $b=n \Rightarrow O(4(n+n))=O(n)$. 


Question 4
There are m sets S_1, S_2,...,S_m.
Each set contains integers in the range [1..n] and m = O(n).
\( n_i = |S_i| = \text{number of elements in } S_i \), \( |S_1| + |S_2| + ... + |S_m| = O(n) \)

Suggest an algorithm for sorting all sets S_1, S_2,...,S_m in O(n) time complexity and O(n) space (memory) complexity.

**Note:** The output is m sorted sets and not one merged sorted set.

Solution:

- If we sort each set using O(nlogn) run time algorithm we get:
  \( T(n) = O(n_1 \log n_1 + n_2 \log n_2 + ... + n_m \log n_m) \leq \log n \cdot O(n_1 + n_2 + ... + n_m) = n \log n \)

- If we sort each set with counting sort we get:
  \( T(n) = O((n_1 + n) + (n_2 + n) + ... + (n_m + n)) = O(n + mn) = O(n^2). \)
  The space complexity is O(n).

The following algorithm runs in O(n) time and space complexity:

1. Add a new variable `set-num` to each element – O(n)
2. Build an array A of all the elements in all the sets - O(n)
3. Sort A using counting-sort on the keys - O(n+n)=O(n)
4. Split A into the original k sets according to the set field – O(n)
   (Counting sort by `set-num`)
**Question 5**

Given \( n \) words in English (assume they are all in lower case). The words are not of the same length. Suggest an algorithm for sorting the words in a lexicographic order in \( O(n) \) time.

**Solution:**

**Reduction:**
Assume you have two problems \( P_1 \) and \( P_2 \) and the algorithm for solving \( P_2 \) is known. A reduction from \( P_1 \) to \( P_2 \) is a way to solve problem \( P_1 \) by using the algorithm for solving \( P_2 \).

In our question the reduction is from lexicographical-sort to radix-sort:

1. **Translate Input of \( P_1 \) to input for \( P_2 \):**
   a. Let \( m \) be the maximal length of word in the input, \( m \) is constant.
   b. Represent each letter as a digit in base 27, i.e., in range \([0..26]\), where \( a=1, b=2, \ldots, z=26 \) - \( O(nm) = O(n) \)
   c. Words with \( k < m \) letters will have succeeding zeros added - \( O(nm) = O(n) \)

2. **Solve \( P_2 \) on the transformed input:**
   a. Execute radix sort where \( d=m \) and \( b=27 \) - \( O(d(n+b)) = O(n) \)

3. **Translate output of \( P_2 \) to output of \( P_1 \):**
   a. Change the numbers sequences back to words - \( O(nm) = O(n) \)

Total run time is \( O(n) \).

**Example:**
Lexicographical-sort input: \{blue, red, green\}
Radix-sort input: \{ (2,12, 21,5,0), (18,5,4,0,0), (7, 18,5,5,14) \}  
Radix-sort output: \{ (2,12, 21,5,0), (7, 18,5,5,14) , (18,5,4,0,0) \}  
Lexicographical-sort output: \{blue, green, red\}
Question 6

Given an array A of n positive integer numbers. All the numbers except for ten are in the range [10, 10n]. Design an algorithm to sort an array A in O(n) time in the worst case.

Solution:

1) Pass on the array and remove 10 numbers that are not in the range into a help array B. O(n) time.
2) Counting sort on the array A. 
   O( n-10 +10n)=O(n) time
3) Some sort on the array B. 
   O(1) time

4) Merge A and B 
   O(n) time