Assignment 08

이름: 최민규(101)

팀원: 문지웅(102), 박규명(102), 정채원(102)

학번: 201812572

**<Q1> 해슁과 비교하여 레드블랙트리의 장점을 2개 이상 제시하라.**

1. 순서가 있는 자료일 경우, 레드블랙트리가 장점을 가집니다. (해쉬는 순서가 없습니다.)
2. 해슁은 페이지 폴트를 일으킬 수 있습니다.
3. 레드블랙트리는 연속된 삽입간에 더 적은 I/O가 발생합니다.
4. 해쉬는 메모리 낭비를 일으킬 수 있습니다.

**<Q2> 각각 m개의 정수와 n개의 정수가 저장된 두 개의 배열 A와 B가 있다. 배열 A가 B의 부분집합인지 아닌지 검사 하는 문제에 대해서 3가지 이상의 방법을 제시하고 각각의 시간복잡도에 대해 논의하라. 해슁을 이용하는 방법을 포함하여야 한다. 입력 정수들은 -2^31에서 2^31 사이의 값들이다.**

**1) 브루트포스(Brute Force) 방법:**

배열 A의 모든 원소가 배열 B에 있는지 확인하는 것입니다. 이를 위해 A의 모든 원소를 순회하면서 각 원소가 배열 B에 있는지 확인합니다.

시간복잡도: O(m\*n)입니다. A와 B의 크기에 따라 선형적으로 증가합니다.

**2)** **정렬과 이진검색(Binary Search) 방법:**

먼저 배열 A, B를 각각 정렬한다. 그 다음 A의 각 원소에 대해 이진 검색을 수행하여 B에 있는지 확인하면 된다.

시간복잡도: **O(mlogn)**입니다. 배열 A와 B를 각각 정렬하는 데 O(mlogm)과 O(nlogn)의 시간이 걸리며, 배열 A의 원소를 한번 이진 검색을 수행하는 데 O(logn)의 시간이 걸립니다. 따라서 m개의 원소를 이진검색을 수행하므로 시간복잡도는 O(mlogn)입니다.

**3)** **해싱(Hashing) 방법:**

우선 배열 B를 해시 테이블에 저장하고 배열 A의 모든 원소에 대해 해시 테이블에서 검색을 수행하면 된다.

시간복잡도: **O(m+n)**입니다. 배열 B를 해시 테이블에 저장하는 데 O(n)의 시간이 걸리며, 배열 A의 모든 원소를 순회하여 검색하는 데 O(m)의 시간이 걸립니다. 이 방법은 배열의 크기가 크더라도 상수 시간 내에 해결할 수 있습니다.

위의 3가지 방법 중에서 가장 효율적인 방법은 해싱 방법입니다.

**<Q3> n개의 영문 대문자가 저장된 배열 X와 하나의 정수 k<=n가 주어진다. 배열에서 길이가 k인 모든 구간에 대해서 그 구간에 포함되는 서로 다른 문자의 개수를 출력하는 문제이다. 예를 들어 X= [A, B, A, C, D, B, C], k=4라면 길이가 4인 첫 번째 구간 [A, B, A, C]에는 3개의 서로 다른 문자가 있고, 두 번째 구간 [B, A, C, D]에는 서로 다른 문자가 4개이다. 다음 구간들인 [A, C, D, B]와 [C, D, B, C]에는 각각 4개와 3개의 서로 다른 문자가 있다. 따라서 출력은 3, 4, 4, 3이다. 이 문제에 대해서 최선의 알고리즘을 제시하고 시간복잡도를 논의하라.**

가장 효율적인 알고리즘은 **슬라이딩 윈도우(Sliding Window) 알고리즘**입니다. 슬라이딩 윈도우 알고리즘은 윈도우를 오른쪽으로 이동하면서 구간 내의 정보를 업데이트하며 해결합니다. 이 문제에서는 구간 내에 있는 서로 다른 문자의 개수를 구해야 하므로, 각 문자의 등장 횟수를 기록하는 해시 테이블을 사용합니다.

이 알고리즘의 시간복잡도는 **O(n)**입니다. 각 문자를 한 번씩만 처리하고, 윈도우의 크기가 k로 고정되어 있기 때문입니다.

**<Q4> n개의 정수가 저장된 배열 A와 하나의 정수 k <= n가 주어진다. 배열에서 길이가 k인 모든 구간에 대해서 그 구 간에 포함되는 서로 다른 정수의 개수를 출력하는 문제이다. 예를 들어 A= [1, 2, 1, 3, 4, 2, 3], k=4라면 길이가 4인 첫 번째 구간 [1, 2, 1, 3]에는 3개의 서로 다른 정수가 있고, 두 번째 구간 [2, 1, 3, 4]에는 서로 다른 정수가 4개이다. 다음 구간들인 [1, 3, 4, 2]와 [3, 4, 2, 3]에는 각각 4개와 3개의 서로 다른 정수가 있다. 따라서 출력은 3, 4, 4, 3이다. 해슁을 사용하는 방법을 포함하여 2가지 이상의 알고리즘을 제시하고 각각의 시간복잡도를 논의하라. 입력 정수들은 -2^31에서 2^31사이의 값들이다.**

1. **브루트포스한 방법**: 길이가 k인 구간을 모두 순회하여 이 과정에서 포함된 정수를 해시맵에 넣어서 해당 구간에 포함된 서로 다른 정수의 개수를 구합니다. 따라서 이 알고리즘의 시간복잡도는 **O(nk)**입니다.
2. **슬라이딩 윈도우 알고리즘**: 먼저 배열의 처음 k개의 원소를 해시맵에 넣고, 해당 구간에 포함된 서로 다른 정수의 개수를 구합니다. 그 다음에는 윈도우를 오른쪽으로 한 칸씩 이동시키면서 구간에 포함된 정수를 추가하고 제거하여 구간 내의 서로 다른 정수의 개수를 구합니다. 따라서 이 알고리즘의 시간복잡도는 **O(n)**입니다.
3. **세그먼트 트리 이용**: 세그먼트 트리를 이용하여 각 구간에 포함된 서로 다른 정수의 개수를 구합니다. 세그먼트 트리는 배열의 구간 합을 빠르게 구하는 자료구조로, 이 알고리즘에서는 각 노드에 해당하는 구간 내의 서로 다른 정수의 개수를 저장합니다. 따라서 이 알고리즘의 시간복잡도는 **O(nlogn)**입니다.