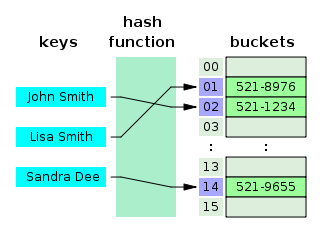
해시 테이블



-key를 input으로 얻은 hash 값을 index로 사용

-O(1)이라는 뛰어난 시간복잡도를 지니나 해시 충돌 문제가 발생

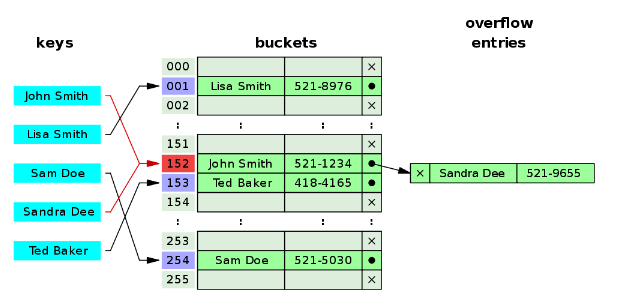
해시 충돌

-2개 이상의 서로 다른 key가 같은 hash값을 지니는 현상

-bucket의 size가 key 전체 size보다 작으므로 필연적으로 발생

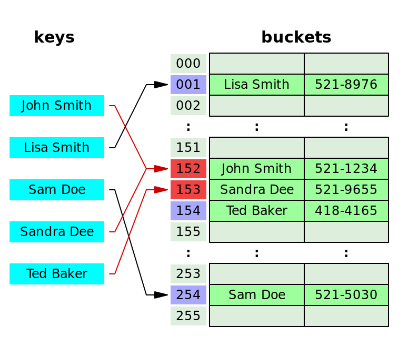
-2가지 해결법, Chaining & Open Addressing

1)Chaining



-한 버켓에 여러 key를 연결리스트로 연결함

2)open addressing



다른 주소에 키를 저장함

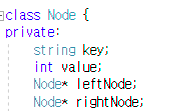


각 방식마다 장단점이 존재하는데, 이번의 경우 key의 개수가 bucket보다 많을 수 있으므로chaining 방식을 사용하여야 한다.

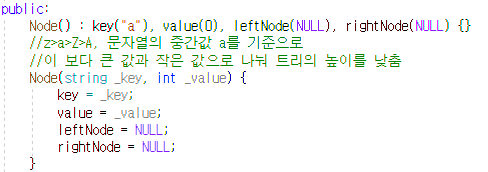
chaining 방식의 문제점

-> 최악의 경우 = 모든 key의 해시 값이 동일한 경우, O(n)이라는 시간 복잡도를 지니게 됨

-> 이를 완화하기 위해, overflow entry를 이진 검색 트리로 구현하였음



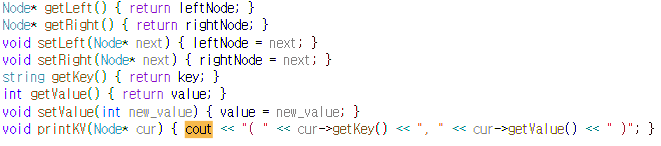
각 node는 key와 value, 그리고 LeftNode, RightNode로 구성됨



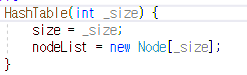
key는 string이며, string 비교 연산시 z>a>Z>A로 계산되므로

전체 문자열의 중간 정도에 위치한 “a” 문자열을 디폴트 값으로 사용,

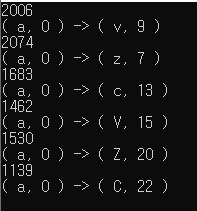
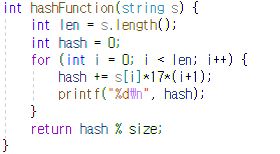
효율적으로 트리의 높이를 낮춤



node의 값을 return하거나 수정, 출력하는 메서드

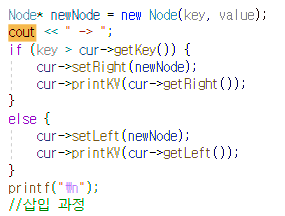
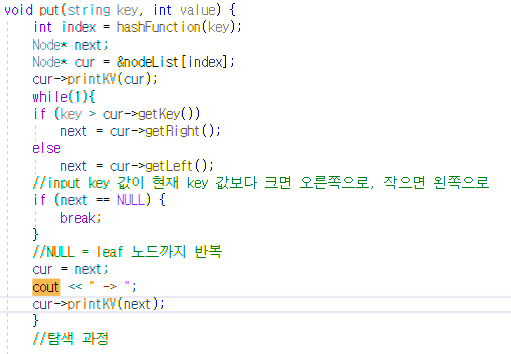


HashTable 생성자, 입력받은 size로 bucket 생성



해시 함수는 digit folding방식을 사용했고,

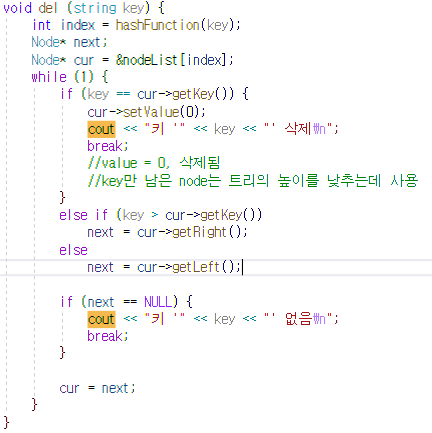
key 문자열이 짧더라도 큰 값이 나올 수 있게 추가적인 값을 곱했다.

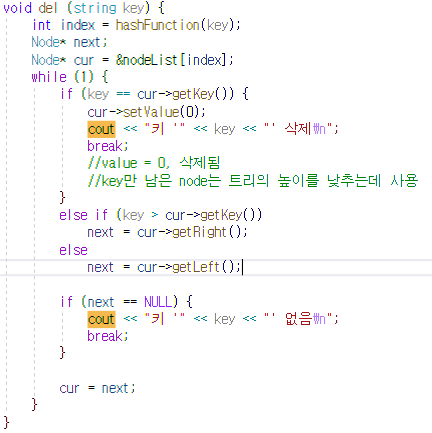


put, 이진 탐색 후 leaf 노드에 도달하면 key를 삽입



get, 이진 탐색 후 input과 동일한 키 값이 있으면 출력

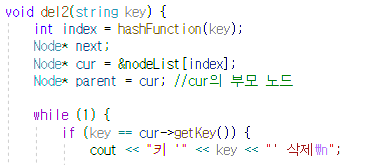




del, 이진 탐색 후 input과 동일한 key 값이 존재하면 value를 0으로 변경

별도의 트리 변경 없이 value만 0으로 바꿈

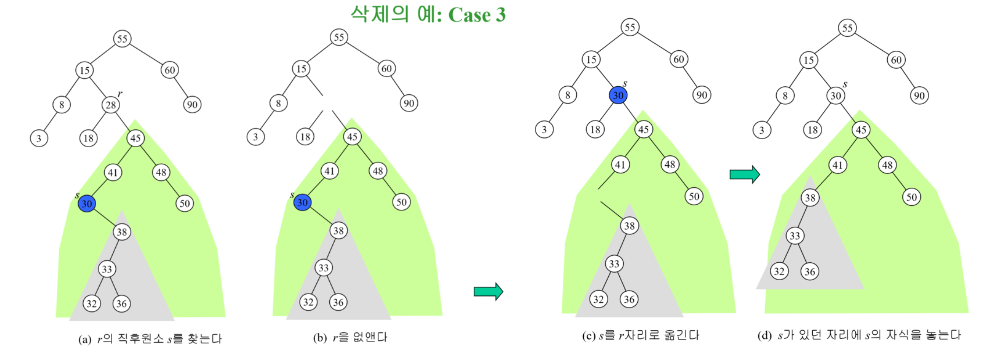
-> 삭제되어야할 노드가 메모리에 남아 공간 증폭이 발생



del2

정석적인 노드 삭제 함수

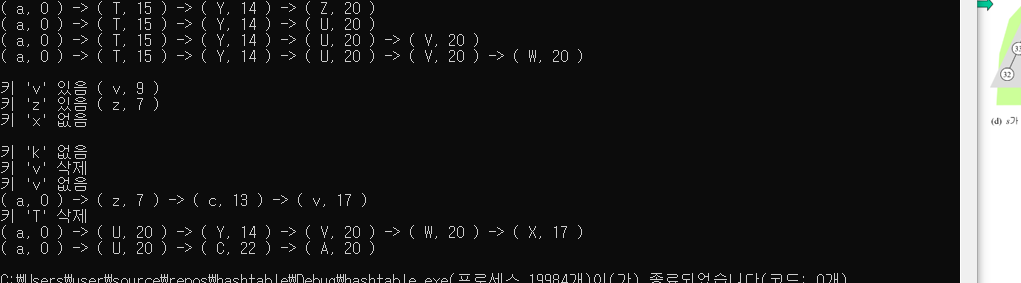
1)삭제하려는 노드의 자식 노드가 2개인 경우

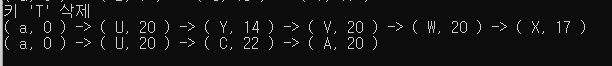


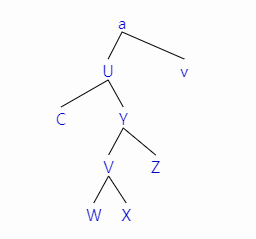
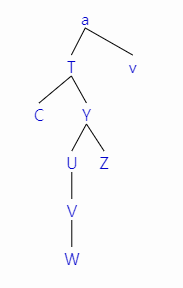
삭제하려는 노드보다 큰 값 중에서 가장 작은 값인 Rmin을 탐색,

노드를 삭제하고 이를 Rmin으로 대체

Rmin의 rightNode는 부모의 leftNode로 연결한다.



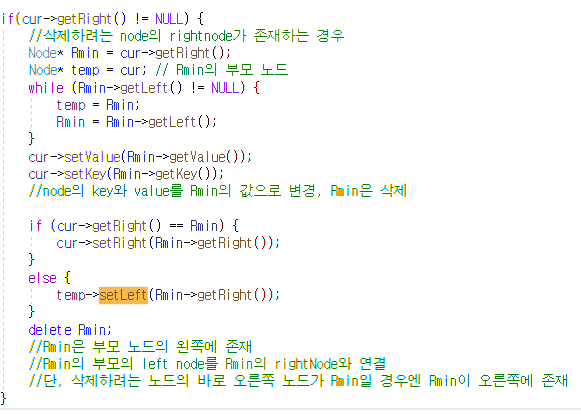
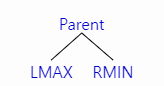




좌측 트리에서 T를 삭제

->T의 Rmin인 U가 T 자리를 대체

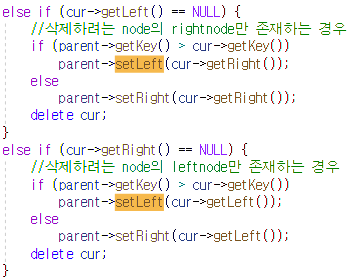
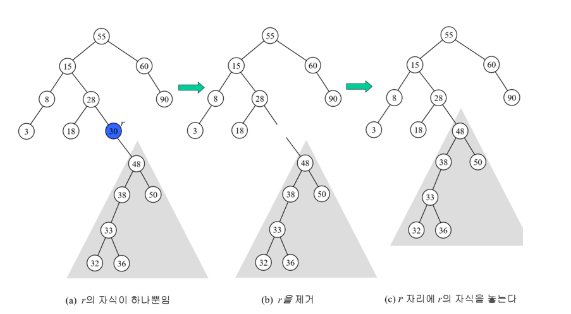
->U의 부모노드 Y와 U의 rightNode인 V를 연결



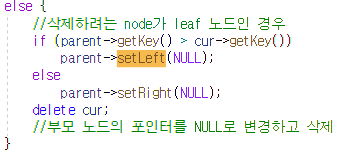
코드 구현

이 때 Rmin은 대부분 부모 노드의 왼쪽에 존재하지만,

삭제하려는 노드의 바로 오른쪽 노드가 Rmin경우 부모 노드의 오른쪽에 존재하므로 예외처리

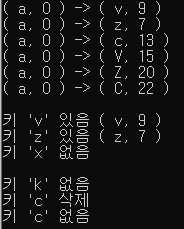
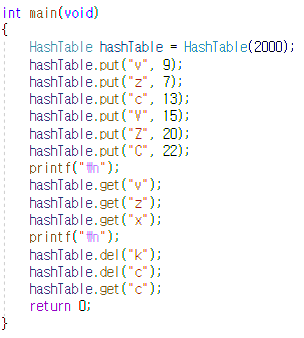


삭제하려는 노드의 자식이 하나뿐일 경우 부모 노드와 자식 노드를 연결하고 삭제하면 된다.



자식이 없을 경우엔 삭제하고 부모 노드의 포인터를 NULL로 바꿔주면 끝

테스트



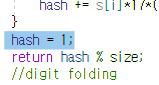
메인 함수에 다음과 같은 key value 쌍을 삽입,

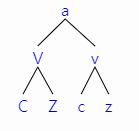
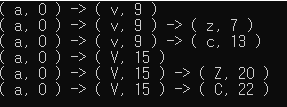
테이블에 존재하는 key와 존재하지 않는 key에 대해 get, del 수행결과

input 시도가 적은 만큼 해시 충돌이 발생하지 않음.

또한 삽입과 삭제가 정상적으로 수행되는 것을 확인

\*(a,0)은 디폴트 root 노드





고의적으로 해시 충돌을 일으키고 key가 이상적으로 삽입 된 경우.

단순 연결 리스트일때보다 평균 시간 복잡도를 낮출 수 있음

이진 탐색 트리는 편향될 수 있으므로 B-tree를 사용하면 더 좋았겠으나,

이번의 메인 과제는 트리가 아닌 해시 테이블이므로

간단히 이진 탐색트리로 구현하였음.

<https://github.com/gillyongs/Algorithm_study/blob/main/1)HashTable/main.cpp>

코드 깃허브 링크