

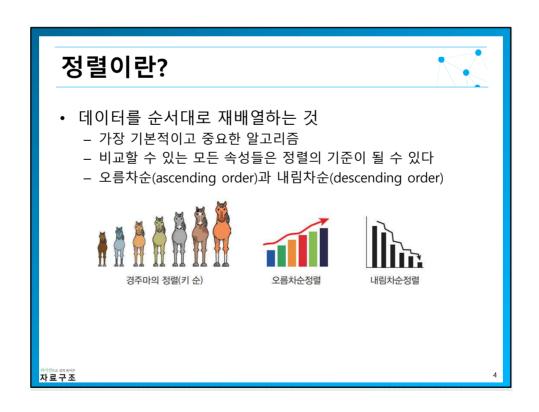
# 7장. 학습 목표

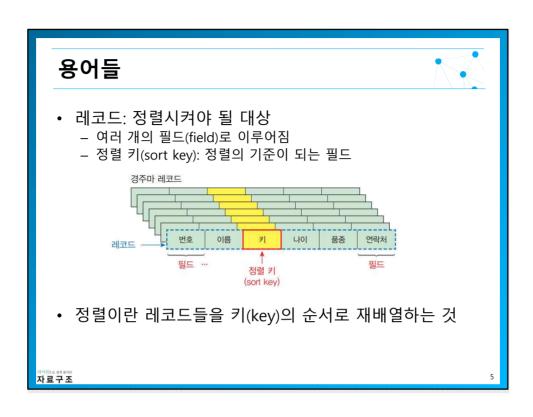


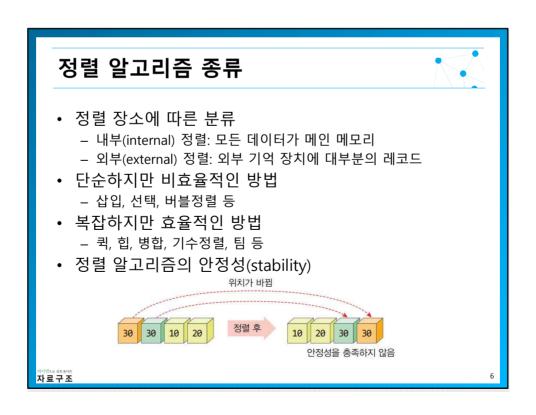
- 정렬의 개념과 간단한 정렬 알고리즘의 동작 원리를 이해한다.
- 정렬을 이용해 집합 관련 연산의 효율을 향상시키는 방법을 이해한다.
- 탐색의 개념과 간단한 탐색 알고리즘의 동작 원리를 이해한다.
- 해싱의 개념과 해시함수, 오버플로의 개념을 이해한다.
- 오버플로 해결방법을 이해하고, 다양한 방법으로 맵을 구현할 수 있다.

파이엔으로 설계 등어준 자 료 구 조

# 







# 7.2 간단한 정렬 알고리즘



- 선택 정렬(selection sort)
- 삽입 정렬(insertion sort)
- 버블 정렬(bubble sort)

파이센으로 4개 등어는 자 료 구 조

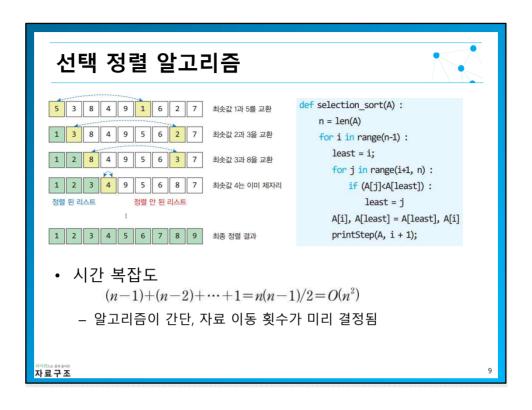
# 선택 정렬(selection sort)

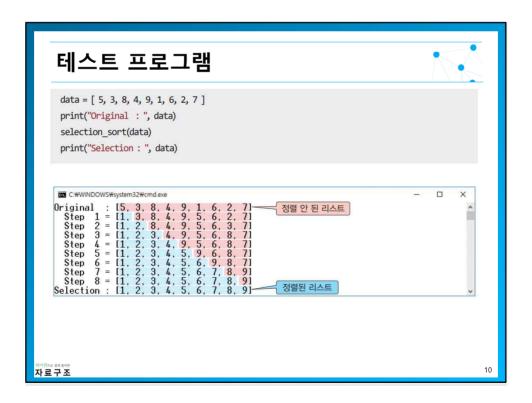


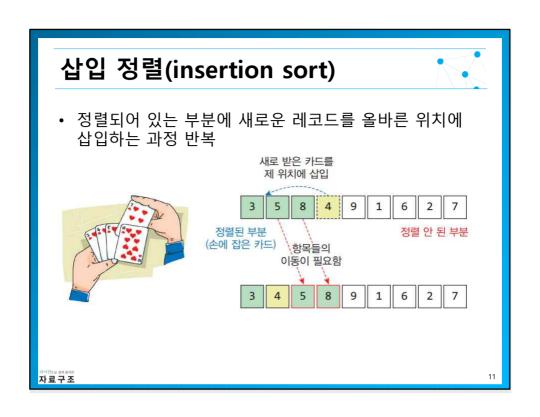
• 오른쪽 리스트에 서 가장 작은 숫자를 선택하여 왼쪽 리 스트의 맨 뒤로 이동하는 작업을 반복

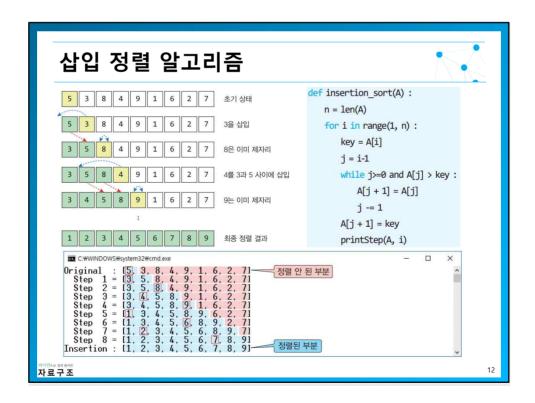
정렬 된(왼쪽) 리스트	정렬 안 된(오른쪽) 리스트	설명 초기상태 1선택 및 이동 2선택 및 이동 3선택 및 이동	
[]	[5,3,8,4,9,1,6,2,7]		
[1]	[5,3,8,4,9,6,2,7]		
[1,2]	[5,3,8,4,9,6,7]		
[1,2,3]	[5,8,4,9,6,7]		
***		4~8 선택 및 이동	
[1,2,3,4,5,6,7,8,9]	[]	9선택 및 이동	

파이젠으로 4개 등어준 **자 료 구 조** 









#### 삽입 정렬 분석



- 복잡도 분석
  - 최선의 경우 O(n): 이미 정렬되어 있는 경우: 비교: n-1 번
  - 최악의 경우 O(n²): 역순으로 정렬되어 있는 경우
    - 모든 단계에서 앞에 놓인 자료 전부 이동

• 
$$\exists l \exists i : \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2} = O(n^2)$$

- 0|\&\varepsilon:\frac{n(n-1)}{2} + 2(n-1) = O(n^2)
- 평균의 경우 O(n²)
- 특징
  - 많은 이동 필요 → 레코드가 큰 경우 불리
  - 안정된 정렬방법
  - 대부분 정렬되어 있으면 매우 효율적

자료구조

12

# 버블 정렬 (bubble sort)



- 기본 전략
  - 인접한 2개의 레코드를 비교하여 순서대로 서로 교환
  - 비교-교환 과정을 리스트의 전체에 수행(스캔)
    - 한번의 스캔이 완료되면 리스트의 오른쪽 끝에 가장 큰 레코드
  - 끝으로 이동한 레코드를 제외하고 다시 스캔 반복

파이쩐으로 쉽게 등어준 자 료 구 조

14



#### 버블정렬 분석



• 비교 횟수(최상, 평균, 최악의 경우 모두 일정)

$$\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2} = O(n^2)$$

- 이동 횟수
  - 역순으로 정렬된 경우(최악): 이동 횟수 = 3 \* 비교 횟수
  - 이미 정렬된 경우(최선의 경우): 이동 횟수 = 0
  - 평균의 경우 : O(n²)
- 레코드의 이동 과다
  - 이동연산은 비교연산 보다 더 많은 시간이 소요됨

파이젠으로 4개 등어준 **자 료 구 조** 

16

### 7.3 정렬 응용: 집합 다시 보기



- 정렬된 리스트를 이용한 집합
- 비교 연산: ea
- 합집합/교집합/차집합
- 복잡도 비교

자료구조

17

#### 정렬 응용: 집합 다시 보기

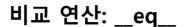


- 3장에서 구현한 집합 자료구조 수정하기
  - 집합의 원소들을 항상 정렬된 순으로 저장
    - 삽입 연산은 더 복잡해 짐
    - 집합의 비교나 합집합, 차집합, 교집합 → 효율적 구현 가능
- 삽입 연산
  - 삽입할 위치를 먼저 찾아야 함.

```
def insert(self, elem): # 정렬된 상태를 유지하면서 elem을 삽입
if elem in self.items: return # 이미 있음
for idx in range(len(self.items)): # loop: n번
if elem < self.items[idx]: # 삽입할 위치 idx를 찾음
self.items.insert(idx, elem) # 그 위치에 삽입
return
self.items.append(elem) # 맨 뒤에 삽입
```

파이쎈으로 4개 등어는 **자 료 구 조** 

10





- 두 집합의 비교 방법
  - 두 집합의 원소의 개수가 같아야 같은 집합이 됨
  - 집합이 정렬되어 있으므로 순서대로 같은 원소를 가져야 함

```
def __eq__( self, setB ): # 두 집합 self, setB가 같은 집합인가?

if self.size() != setB.size() : # 원소의 개수가 같아야 함

return False

for idx in range(len(self.items)): # loop: n번

if self.items[idx] != setB.items[idx] : # 원소별로 같은지 검사

return True
```

• 시간 복잡도:*O*(*n*<sup>2</sup>) → *O*(*n*)으로 개선

자료구조

19

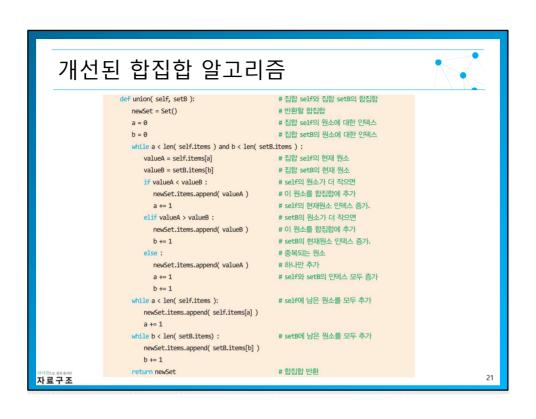
#### 합집합/교집합/차집합



- 합집합 연산 방법
  - 가장 작은 원소들부터 비교하여 더 작은 원소를 새로운 집합에 넣고 그 집합의 인덱스를 증가시킴.
  - 만약 두 집합의 현재 원소가 같으면 하나만을 넣음. 인덱스는 모두 증가시킴.
  - 한쪽 집합이 모두 처리되면 나머지 집합의 남은 모든 원소를 순 서대로 새 집합에 넣음
- 시간 복잡도: *O*(*n*<sup>2</sup>) → *O*(*n*)으로 개선
- 교집합과 차집합도 동일한 방법 적용 가능

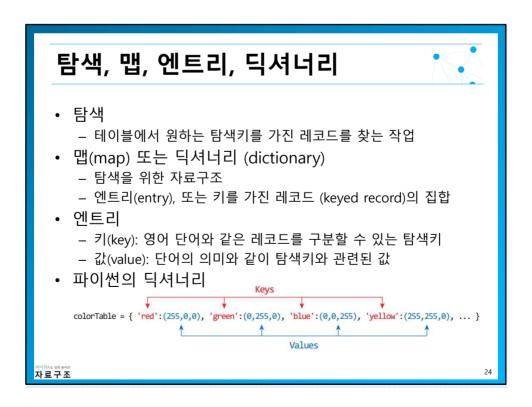
파이쩐으로 쉽게 등어준 자 료 구 조

20



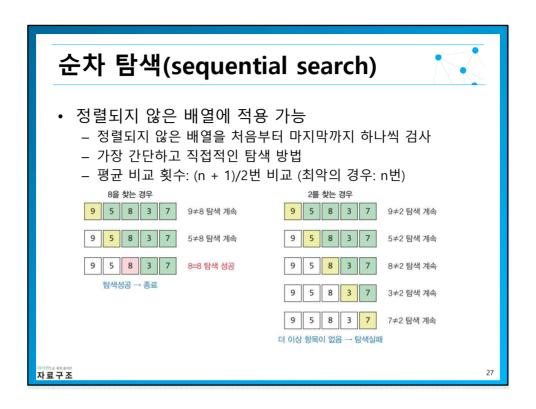


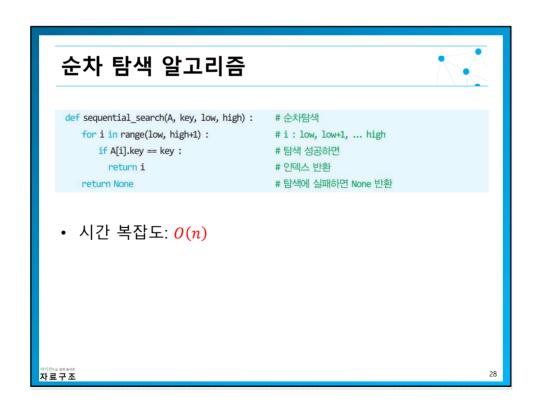
# 7.4 탐색과 맵 구조 • 탐색, 맵, 엔트리, 딕셔너리 • 맵 ADT











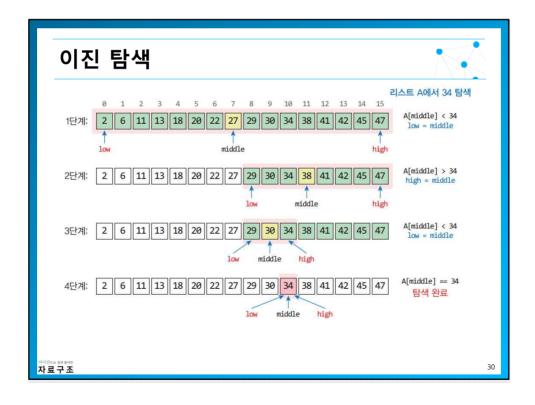
# 이진 탐색(binary search)



- 정렬된 배열의 탐색에 적합
  - 배열의 중앙에 있는 값을 조사하여 찾고자 하는 항목이 왼쪽
     또는 오른쪽 부분 배열에 있는지를 알아내어 탐색의 범위를 반으로 줄여가며 탐색 진행
  - 예) 사전에서 단어 찾기
- (예) 10억 명중에서 특정한 이름 탐색
  - 이진탐색 : 단지 30번의 비교 필요
  - 순차 탐색 : 평균 5억 번의 비교 필요

자료구조

29



```
OIO 탐색 알고리즘

def binary_search(A, key, low, high):
    if (low <= high): # 항목들이 남아 있으면(종료 조건)
    middle = (low + high) // 2 # 정수 나눗셈 //에 주의할 것.
    if key == A[middle].key: # 탐색 성공
        return middle
    elif (key<A[middle].key): # 왼쪽 부분리스트 탐색
        return binary_search(A, key, low, middle - 1)
    else: # 오른쪽 부분리스트 탐색
        return binary_search(A, key, middle + 1, high)
    return None # 탐색 실패

• 시간 복잡도: O(bg n)
• 반복으로 구현 가능
```

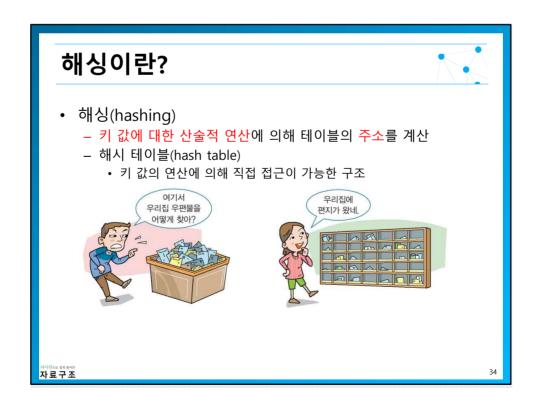
# 7.6 고급 탐색 구조: 해싱

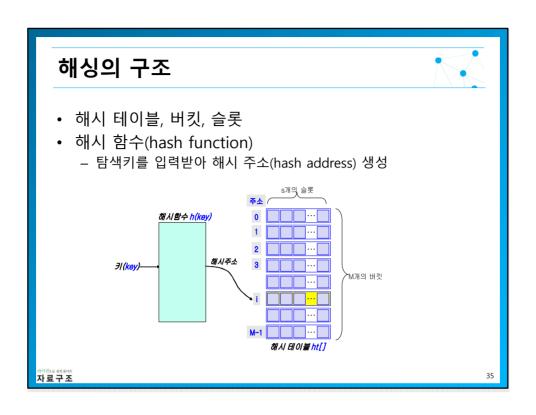


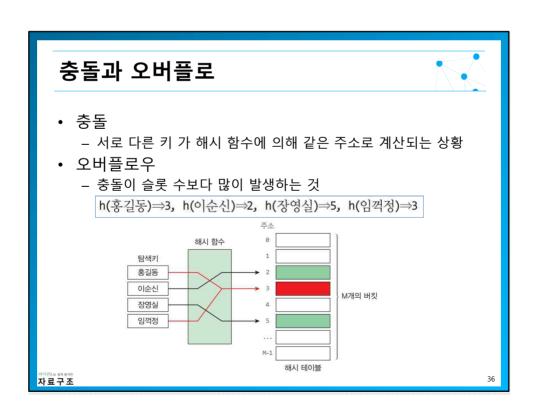
- 해싱이란?
- 선형 조사에 의한 오버플로 처리
- 체이닝(chaining)에 의한 오버플로 처리
- 해시 함수
- 탐색 방법들의 성능 비교

파이센으로 4개 등어는 자 료 구 조

33

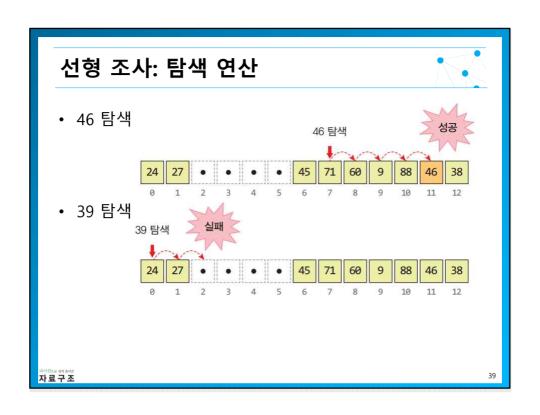


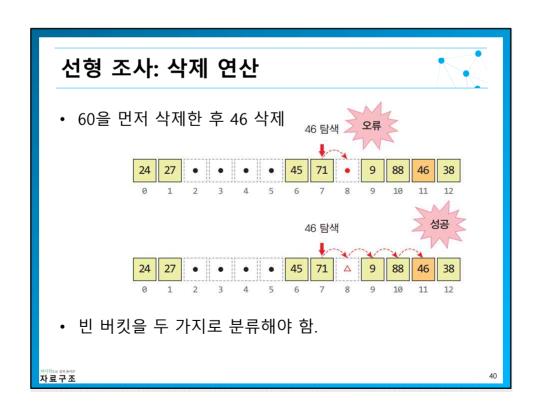












#### 선형 조사 군집화 완화 방법



• 이차 조사법(quadratic probing)

 $(h(k)+i^*i)\% M$  for  $i=0,1,\dots,M-1$ 

- 이중 해싱법(double hashing)
  - 재해싱(rehashing)
  - 충돌이 발생하면, 다른 해시 함수를 이용해 다음 위치 계산

파이벤으로 설계 등어준 자 료 구 조 41

#### 체이닝에 의한 오버플로 처리



- 하나의 버킷에 여러 개의 레코드를 저장할 수 있도록 하는 방법
- 예) h(k)=k%7 을 이용해 8, 1, 9, 6, 13 을 삽입

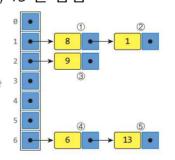
① 8 저장 : h(8) = 8 % 7 = 1 ⇒ 저장

② 1 저장 : h(1) = 1% 7 = 1  $\Rightarrow$  충돌  $\Rightarrow$  새로운 노드 생성 및 저장

③ 9 저장 : h(9) = 9 % 7 = 2  $\Rightarrow$  저장

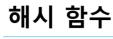
④ 6 저장 : h(6) = 6 % 7 = 6  $\Rightarrow$  저장

⑤ 13저장 :  $h(13) = 13 \% 7 = 6 \Rightarrow 충돌 <math>\Rightarrow$  새로운 노드 생성 및 저장



파이쩐으로 쉽게 등어준 자 료 구 조

및 보이 휴아트 ! 그 ㅈ





- 좋은 해시 함수의 조건
  - \_ 충돌이 적어야 한다
  - 함수 값이 테이블의 주소 영역 내에서 고르게 분포되어야 한다
  - 계산이 빨라야 한다
- 제산 함수
  - $h(k)=k \mod M$
  - 해시 테이블의 크기 M은 소수(prime number) 선택
- 폴딩 함수



자료구조

#### 해시 함수



- 중간 제곱 함수
  - 탐색키를 제곱한 다음, 중간의 몇 비트를 취해서 해시 주소 생성
- 비트 추출 함수
  - 키를 이진수로 간주. 임의의 위치의 k개의 비트를 사용
- 숫자 분석 방법
  - 키에서 편중되지 않는 수들을 테이블의 크기에 적합하게 조합
- 탐색키가 문자열인 경우

```
def hashFn(key) :
        sum = 0
         for c in key:
                               # 문자열의 모든 문자에 대해
            sum = sum + ord(c) # 그 문자의 아스키 코드 값을 sum에 더함
        return sum % M
자료구조
```

# 탐색 방법들의 성능 비교



• 해싱의 적재 밀도(loading density) 또는 적재 비율 - 저장되는 항목의 개수 n과 해시 테이블의 크기 M의 비율

 $lpha = rac{ ext{저장된 항목의 개수}}{ ext{해성테이블의 버킷의 개수}} = rac{n}{M}$ 

• 다양한 탐색 방법의 성능 비교

탐색방법		탐색	삽입	삭제
순차탐색		O(n)	O(1)	O(n)
이진탐색		$O(\log_2 n)$	O(n)	O(n)
이진탐색트리 -	균형트리	$O(\log_2 n)$	$O(\log_2 n)$	$O(\log_2 n)$
	경사트리	O(n)	O(n)	O(n)
해싱	최선의 경우	O(1)	O(1)	O(1)
	최악의 경우	O(n)	O(n)	O(n)

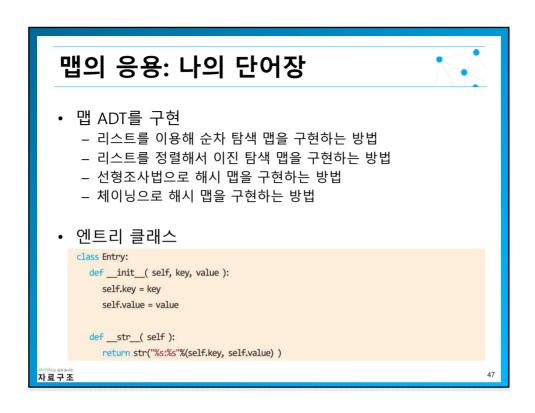
파이벤으로 4개 등어온 자 료 구 조 45

# 7.7 맵의 응용: 나의 단어장



- 리스트를 이용한 순차탐색 맵
- 체이닝을 이용한 해시 맵
- 파이썬의 딕셔너리를 이용한 구현

파이쩐으로 쉽게 등어준 자 료 구 조 46



```
리스트를 이용한 순차탐색 맵
    class SequentialMap:
                                               # 순차탐색 맵
      def __init__( self ):
         self.table = []
                                               # 맵의 레코드 테이블
       def insert(self, key, value) :
                                               # 삽입 연산
         self.table.append(Entry(key, value))
                                               # 리스트의 맨 뒤에 추가
       def search(self, key) :
                                               # 순차 탐색 연산
          pos = sequential_search(self.table, key, 0, self.size()-1)
          if pos is not None : return self.table[pos]
          else : return None
       def delete(self, key):
                                           # 삭제 연산: 항목 위치를 찾아 pop
          for i in range(self.size()):
                                           # 삭제할 위치를 먼저 찾고
            if self.table[i].key == key :
              self.table.pop(i)
                                           # 리스트의 pop으로 삭제
자료구조
```

