제10장 검색

1. 개요

검색(탐색)은 기억된 자료 중에서 특정 조건을 만족하는 자료를 찾는 것이다.

검색 방법의 종류는 다음과 같이 분류할 수 있다.

- 선형검색(linear search)
- 제어검색
  - 이진검색(binary search)
    - 피보나치검색(fibonacci search)
    - 보간탐색(interpolation search)
- 블록검색(block search)
- 트리검색(tree search)
- 해싱(hashing) → 비조사 비교법(키값의 계수적인 성질을 이용)

• **제어검색**은 **정렬된 리스트**에서만 수행할 수 있다.

◈ 기본 용어 정리

• 키(key) : 레코드를 구별하기 위해 사용되는 필드(각 필드는 키가 될 수 있다)

• 주요키(primary key) : 각 레코드를 완전히 구분할 수 있는 키(통상 키라고도 함)

전 화 번 호 부

이름	전화번호	주소
놀부	825-2222	서울 동작구 노량진 1동
흥부	825-3333	서울 동작구 노량진 3동
길동	825-7777	서울 동작구 노량진 2동
놀부	825-5555	서울 동작구 노량진 4동

전화번호는 Primary key가 될 수도 있다.

이름은 Primary key가 될 수 없다.

• 내부검색 ~ 주기억공간에 수록된 자료에서 특정 자료를 검색
• 외부검색 ~ 보조기억공간에 수록된 자료에서 특정 자료를 검색

• 검색장(search length) : 특정 레코드를 찾기 위한 비교회수
• 평균검색장(average search length) : 검색 단위 동작의 평균비교회수

• **평균검색장(L)** = $\frac{\text{각 자료를 검색하는 비교회수의 총합}}{\text{검색을 요구한 비교회수}}$

[예제] 접근빈도수가 있는 목록에서 평균검색장(L)

인덱스	레코드	빈도수
1	R1	30
2	R2	40
3	R3	10
4	R4	20

$L = \frac{1\times30+2\times40+3\times10+4\times20}{30+40+10+20} = 2.2$(회)

• 빈도수가 많은 레코드 순으로 목록이 작성될수록 평균검색장은 짧아진다.

기출문제 분석

1. 다음 검색 알고리즘에서 Data를 검색하는 방법에서 성격이 다른 하나는? [2002년 국가 7급]

① 이진검색 ② 트리검색
③ 보간탐색 ④ 해싱

♤ **해싱**

• 해싱은 비조사 비교법이다. 즉, 키의 **계수적인 성질**을 이용한다.

정답 : ④

2. 다음은 검색 및 정렬 알고리즘에 관한 특징을 나열한 것이다. 잘못 기술한 것은? [2015년 국회 9급]

① 순차검색(sequential search) : 모든 레코드를 처음부터 탐색한다.
② 해쉬(hash) : 레코드 양과 관계없이 검색시간이 일정하다.
③ 트리검색(tree search) : 이진트리를 구성하여 실행한다.
④ 이분검색(binary search) : 자료가 정렬(sort)되어 있어야 가능하다.
⑤ 삽입정렬(insertion sort) : 최악의 경우 $O(n^2)$ 시간을 필요로 한다.

♤ **검색 및 정렬 알고리즘**

---

② 해쉬(hash) : 레코드 양과 관계없이 검색시간이 일정하다.**(×)**
- 오버플로가 발생되면, 검색시간은 레코드 양이 많을수록 길어질 수 있다.
  → 충돌이 발생되면, 검색시간은 길어질 수밖에 없다.
  → 충돌이 발생되면, 최악의 경우 검색시간은 선형시간인 O(n)이 된다.
- 오버플로가 발생되지 않으면, 검색시간은 레코드 양과 관계없이 일정하다.
  → 충돌이 발생되지 않으면, 검색시간은 해시표 크기와는 큰 관계가 없다.
  → 충돌이 발생되지 않으면, 검색시간은 O(1)이 된다. 일정하다는 의미이다.

③ **트리검색**(tree search) : **이진트리**를 구성하여 실행한다.**(×)**
  → **트리검색**은 **이진탐색트리**를 구성하여 실행한다.
  → 이진탐색트리는 이진트리의 특수구조이기 때문이다.
  → 모든 이진트리가 이진탐색트리가 될 수 없다.

// 이의 신청한 문제
- 이 문제의 정답은 처음에 ②로 발표되었다.
- 이의 신청한 결과 정답은 ②, ③으로 수정 발표되었다.
- 문제를 정확하게 출제하지 않았기에, 드물게 복수 정답이 인정된 경우이다.
- 복수 정답이 인정되어 합격한 분도 있었다.

정답 : ②, ③