

[DB 9강]-자료 구조(정렬, 검색, 해싱)

4. 버블정렬(Bubble Sort) ★☆☆☆☆

-> 수면 위로 올라가는 물방울 (인접 데이터 비교)

점수 100 70 90 80 90> 70 90 80 90 100 (Pass 1)
> 70 80 90 90 100 (Pass 2)
> 70 80 90 90 100 (Pass 3)
> 70 80 90 90 100 (Pass 4)

Pass 1 일 때
 교 환 100 70 90 80 90
 교 환 70 100 90 80 90
 교 환 70 90 100 80 90
 교 환 70 90 80 100 90
 교 환 70 90 80 90 100

Pass 3 일 때
 변화 없음 70 80 90 90 100
 변화 없음 70 80 90 90 100

Pass 2 일 때
 변화 없음 70 90 80 90 100
 교 환 70 90 80 90 100
 변화 없음 70 80 90 90 100
 3

Pass 4 일 때
 변화 없음 70 80 90 90 100
 변화 없음 70 80 90 90 100



[DB 9강]-자료 구조(정렬, 검색, 해싱)

☆☆☆☆☆

5. 삽입 정렬(insertion Sort)

-> 성적순으로 교실 자리 배치 하기

점수 100 70 90 80 90> 70 100 90 80 90 (Pass 1)
> 70 90 100 80 90 (Pass 2)
> 70 80 90 100 90 (Pass 3)
> 70 80 90 90 100 (Pass 4)

[70 삽입하기]

Pass 1 일 때
 이 동 100 70 90 80 90
 삽 입 70 100 90 80 90

[80 삽입하기]

Pass 3 일 때
 이 동 70 90 100 80 90
 이 동 70 90 100 90
 변화없음 70 90 100 90
 삽 입 70 80 90 100 90

[90 삽입하기]

Pass 2 일 때
 이 동 70 100 90 80 90
 변화없음 70 100 80 90
 삽 입 70 90 100 80 90

[90 삽입하기]

Pass 4 일 때
 이 동 70 80 90 100 90
 변화없음 70 80 90 100
 변화없음 70 80 90 100
 변화없음 70 80 90 100
 삽 입 70 80 90 90 100

* Pass 1에서 '두 번째 값'을 Key 로 선택한다.



[DB 9강]-자료 구조(정렬, 검색, 해싱)



6. 2-Way 병합 정렬

-> 두 개씩 묶어서 정렬

71, 2, 38, 5, 7, 61, 11, 26, 53, 42

1회전 : (71, 2), (38, 5), (7, 61), (11, 26), (53, 42)
(2, 71), (5, 38), (7, 61), (11, 26), (42, 53)

2회전 : ((2, 71), (5, 38)), ((7, 61), (11, 26)), (42, 53)
(2, 5, 38, 71), (7, 11, 26, 61), (42, 53)

3회전 : ((2, 5, 38, 71), (7, 11, 26, 61)), (42, 53)
(2, 5, 7, 11, 26, 38, 61, 71), (42, 53)

4회전 : ((2, 5, 7, 11, 26, 38, 61, 71), (42, 53))
2, 5, 7, 11, 26, 38, 42, 53, 61, 71

[DB 9강]-자료 구조(정렬, 검색, 해싱)

정렬해야 할 파일이(5, 1, 4, 3, 8, 2)인 6개의 키 값을 첫번째 단계에서 3회 수행한 결과가 다음과 같을 때, 어떤 정렬 기법을 사용하였는가? <답은 나>

초기 : 5, 1, 4, 3, 8, 2	1회 : 1, 5, 4, 3, 8, 2
2회 : 1, 4, 5, 3, 8, 2	3회 : 1, 4, 3, 5, 8, 2

가. 삽입 정렬 나. 버블 정렬 다. 합병 정렬 라. 힙 정렬

=> 초기에서 1회전 후에 상태는 1,4,3,5,2,8 입니다. (문제에서는 단계라고 했으므로 단계=회전 으로 생각하세요.)

1회전을 좀 더 상세히 살펴 보겠습니다. (문제에서는 회라고 했습니다.)

초기 : 5, 1, 4, 3, 8, 2 -> 1회 : 1, 5, 4, 3, 8, 2

초기 5,1 을 비교해서 큰 값을 우측으로 이동합니다. 결과 : 1,5

1회 : 1, 5, 4, 3, 8, 2 -> 2회 : 1, 4, 5, 3, 8, 2

1회 5,4 를 비교해서 큰 값을 우측으로 이동합니다. 결과 : 4,5

2회 : 1, 4, 5, 3, 8, 2 -> 3회 : 1, 4, 3, 5, 8, 2

2회 5,3 을 비교해서 큰 값을 우측으로 이동합니다. 결과 : 3,5

[DB 9강]-자료 구조(정렬, 검색, 해싱)

1. 검색

: 원하는 데이터를 찾을 때.

1) 선형 검색 (순차검색)

- 모든 레코드들을 대상으로 순차적 검색, 자료가 정렬되지 않을 때, 처리 속도가 느리다

2) 이분(이진) 검색 ★★☆☆☆☆

- 자료가 정렬되어 있어야 함, 중간 값을 비교 검색, 많은 레코드 검색 시 효율적

10	15	31	45	62	63	74	87	91	92
----	----	----	----	----	----	----	----	----	----

4) 보간 검색 : 찾고자 하는 레코드 키가 있음직한 위치를 추정

5) 트리 검색, 블럭 검색, 피보나치 검색

[DB 9강]-자료 구조(정렬, 검색, 해싱)

2. 해싱(hashing) 검색 ★★☆☆☆☆

: **해싱 함수**를 이용하여 자료를 검색하는 방법. 데이터를 **해시 테이블**이라는 배열에 저장하고, 해싱 함수를 이용하여 데이터가 위치한 곳의 주소를 찾기 때문에 신속하게 원하는 자료를 검색할 수 있는 **키-주소 변환 방법**

1) 특징

- DAM(직접접근, Direct Access Method) 파일을 구성할 때 해싱이 사용
- 삽입, 삭제 작업의 빈도가 많을 때 유리한 방식
- 검색은 가장 빠르지만 기억공간의 낭비 발생

2) 용어 정리

- 해싱 함수 : 해시 테이블의 주소를 생성해 내는 함수
- 해시 테이블 : 해싱 함수에 의하여 참조되는 테이블
- 버킷 (bucket) : 하나의 주소를 갖는 파일의 한 구역
- 슬롯 (slot) : n개의 슬롯이 모여 하나의 버킷을 형성
- 충돌 (collision) : 서로 다른 2개 이상의 레코드가 같은 주소를 갖는 현상
- 시노임 (Synonym) : 같은 주소를 갖는 레코드의 집합
- 오버플로 : 버킷 내에 기억 공간이 없는 현상

8 7 6 9 4 14 2 15 19

7로 나눈 후
나머지 계산

0	1	2	3	4	5	6

해싱 함수

해시 테이블

[DB 9강]-자료 구조(정렬, 검색, 해싱)

1. 정렬(Sort) 알고리즘의 선택 시 고려 사항으로 거리가 먼 것은?

- 가. 증가 데이터의 배열 상태
- 나. 키 값들의 분포 상태
- 다. 소요공간 및 작업시간
- 라. 정렬에 필요한 기억공간의 크기

2. 주기억장치 내에서 정렬하는 방법에 해당하는 것은?

- 가. Cascade Merge Sort 나. Heap Sort
- 다. Balanced Merge Sort 라. Polyphase Merge Sort

3. 외부 정렬(External Sort)에 해당하지 않는 것은?

- 가. Balanced Sort 나. Cascade Sort
- 다. Insertion Sort 라. Polyphase Sort

4. 인접한 데이터를 비교하면서 그 크기에 따라 데이터의 위치를 바꾸어 정렬하는 방법은?

- 가. 퀵 정렬(Quick Sort) 나. 힙 정렬(Heap Sort)
- 다. 셸 정렬(Shell Sort) 라. 버블 정렬(Bubble Sort)

5. 자료가 아래와 같을 때, 삽입(Insertion) 정렬 방법을 적용하여 오름차순으로 정렬할 경우 Pass 1을 수행한 결과는?

20, 19, 14, 16, 18

- 가. 19, 20, 14, 16, 18 나. 14, 20, 19, 16, 18
- 다. 14, 19, 20, 16, 18 라. 20, 14, 19, 16, 18

6. 자료가 아래와 같이 주어졌을 때, 선택 정렬(Selection Sort)을 적용하여 오름차순으로 정렬할 경우 Pass 2를 진행한 후의 정렬된 값으로 옳은 것은?

9, 4, 5, 11, 8

- 가. 4, 5, 9, 8, 11
- 나. 4, 5, 9, 11, 8
- 다. 4, 5, 8, 11, 9
- 라. 4, 5, 8, 9, 11

[정답] 1.가 2.나 3.다 4.라 5.가 6.나

[DB 9강]-자료 구조(정렬, 검색, 해싱)

7. 레코드의 많은 자료 이동을 없애고 하나의 파일을 부분적으로 나누어가면서 정렬하는 방법으로 키를 기준으로 작은 값은 왼쪽에 큰 값은 오른쪽에 모이도록 서로 교환시키는 부분 교환 정렬법은?

- 가. 퀵 정렬(Quick Sorting)
- 나. 셸 정렬(Shell Sorting)
- 다. 삽입 정렬(Insertion Sorting)
- 라. 선택 정렬(Selection Sorting)

8. 이진 트리의 레코드 R=(88, 74, 63, 55, 37, 25, 33, 19, 26, 14, 9)에 대하여 힙(Heap) 정렬을 만들 때, 37의 왼쪽과 오른쪽의 자노드(Child Node)의 값은?

- 가. 55, 25 나. 63, 33
- 다. 33, 19 라. 14, 9

9. 연산 시간이 최악과 평균의 경우, 모두 $O(n \log n)$ 으로 빠른 속도를 갖는 정렬 방식은?

- 가. 퀵 정렬(quick sort) 나. 버블 정렬(bubble sort)
- 다. 힙 정렬(heap sort) 라. 선택 정렬(selection sort)

10. 다음은 어떠한 정렬 방법을 설명한 것인가?

- 두 개의 키들을 한 쌍으로 하여 각 쌍에 대하여 순서를 정한다.
- 순서대로 정렬된 각 쌍의 키들을 합병하여 하나의 정렬된 서브리스트로 만든다.
- 위 과정에서 정렬된 서브리스트들을 하나의 정렬된 파일이 될 때까지 반복한다.

- 가. 2-Way 합병 정렬
- 나. 퀵 정렬
- 다. 기수 정렬
- 라. 버블 정렬

11. 정렬 알고리즘 선택 시 고려하여야 할 사항으로 거리가 먼 것은?

- 가. 데이터의 양
- 나. 초기 데이터의 배열상태
- 다. 키 값들의 분포상태
- 라. 운영체제의 종류

[정답] 7.가 8.라 9.다 10.가 11.라

[DB 9강]-자료 구조(정렬, 검색, 해싱)

12. 다음과 같이 레코드가 구성되어 있을 때, 이진 검색 방법으로 14를 찾을 경우 비교되는 횟수는?

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

가. 2번 나. 3번 다. 4번 라. 5번

13. 아래 보기의 자료에서 이진 탐색(Binary Search)을 적용할 경우 m을 찾기 위한 비교 횟수는?

a b c d e f g h i j k l m n o

가. 3 나. 4 다. 5 라. 6

14. 해시(Hash) 함수와 밀접한 관계가 있는 파일은?

가. ISAM 파일 나. VSAM 파일
다. DAM 파일 라. 링 파일

15. 해싱 함수의 값을 구한 결과 키 k1, k2가 같은 값을 가질 때, 이들 키 k1, k2의 집합을 무엇이라 하는가?

가. Mapping 나. Folding 다. Synonym 라. Chaining

[정답] 12.나 13.나 14.다 15.다 16.나 17.라
18.다 19.가

11

16. 키 값을 여러 부분으로 분류하여 분류한 부분을 더하거나 XOR하여 주소를 계산하는 해싱 함수의 종류는?

가. 제산(Divide)함수
나. 접지(Folding)함수
다. 중간제곱(Mid-Square)함수
라. 숫자 분석 함수

17. 버킷(bucket)과 가장 관련이 깊은 것은?

가. SAM 나. ISAM 다. B-Tree 라. Hashing

18. 해싱함수(Hashing Function)의 종류가 아닌 것은?

가. 제곱(Mid-Square) 방법
나. 숫자분석(Digit Analysis) 방법
다. 체인(Chain) 방법
라. 제산(Division) 방법

19. 검색 방법 중 속도는 가장 빠르지만 충돌현상 시 오버플로 해결의 부담이 과중되며, 많은 기억공간을 요구하는 탐색 방법은?

가. 해싱(hashing)
나. 블록 탐색(block search)
다. 순차 탐색(sequential search)
라. 이진 탐색(binary search)



[DB 9강]-자료 구조(정렬, 검색, 해싱)

20. 해싱(hashing)에서 서로 다른 키(key)가 같은 홈주소(home address)를 가지는 경우를 무엇이라 하는가?

가. 동의어(synonym)
나. 재귀(recursion)
다. 충돌(collision)
라. 버킷(bucket)

22. 다음의 자료를 삽입(insert) 정렬 기법을 사용하여 오름차순으로 정렬할 경우 PASS 2 의 결과는?

64 28 33 76 55 12 43

가. 28 64 33 76 55 12 43
나. 12 28 33 55 64 76 43
다. 28 33 64 76 55 12 43
라. 12 64 28 33 76 55 43

21. 해싱(Hashing)에 관한 설명으로 옳지 않은 것은?

가. 버킷(bucket)이란 하나의 주소를 갖는 파일의 한 구역을 의미하며, 버킷의 크기는 같은 주소에 포함될 수 있는 레코드 수를 의미한다.
나. 슬롯(slot)이란 한 개의 레코드를 저장할 수 있는 공간으로 n개의 슬롯이 모여 하나의 버킷을 형성한다.
다. 충돌(collision)이란 레코드를 삽입할 때 2개의 상이한 레코드가 똑같은 버킷으로 해싱되는 것을 의미한다.
라. 해싱은 충돌(collision)이 발생하면 항상 오버플로우(overflow)가 발생한다.

[정답] 20.다 21.라 22.다

12



[DB 9강]-자료 구조(정렬, 검색, 해싱)

[산-08년3월]

23. 해싱 함수 기법에서 키 값을 양의 정수인 소수로 나누어 나머지를 홀 주소로 취하는 방법을 무엇이라고 하는가?

- 가. 폴딩(Folding)법 나. 제곱(Mid-Square)법
다. 제산(Division)법 라. 기수(Radix)변환법

[기-08년5월][산-08년3월]

24. 해싱 기법에서 동일한 홀 주소로 인하여 충돌이 일어난 레코드들의 집합을 무엇이라고 하는가?

- 가. Synonym 나. Collision
다. Bucket 라. Overflow

[해싱함수 종류]

- 기수(Radix)법 : 다른 진법으로 변환
- folding(중첩)법 : 더하거나 XOR
- 계수 분석법 : 숫자의 분포 조사
- 제산(Division)법 : 나누어 나머지
- 제곱법 : 제곱

[오버플로우 해결 방법]

- 개방주소법 : 다음 빈 버킷에 저장
- 폐쇄주소법 : 별도의 공간에 저장
- 재해싱법 : 새로운 해싱함수를 적용

[정답] 23.다 24.가