

<용어체크>

해시 충돌

해시 함수를 통해 만들어진 해시 주소가 중복되면 데이터 값이 충돌하는데 이러한 충돌을 피하는 방법은 충돌이 적은 좋은 해시 함수를 사용하는 것이며 충돌이 발생했을때 대표적인 충돌 해결 방법에는 체이닝, 개방 주소법이 있다.

체이닝

같은 주소로 해싱되는 원소를 모두 하나의 연결 리스트로 관리하는 방법이며 연결 리스트만 사용하면 되므로 복잡한 계산식을 사용할 필요가 개방 주소 방법에 비해 적다.

개방 주소 방법

해시 충돌이 일어나면 다른 버킷에 데이터를 삽입하는 방식인데 빈자리가 생길 때까지 해시값을 계속 만들어내므로 체이닝 처럼 포인터가 필요 없고 지정된 메모리 외 추가적인 저장 공간도 필요 없어 삽입, 삭제 시 오버헤드가 적다.

<학습내용>

해시 테이블의 충돌 해결

체이닝

개방 주소 방법

<학습목표>

해시 테이블의 충돌 해결 방법에 대해 설명할 수 있다.

체이닝 방법에 대해 이해하고 장단점을 설명할 수 있다.

개방 주소 방법에 대해 이해하고 장단점을 설명할 수 있다.

Q. 해싱은 해시 테이블과 해시 함수로 구성되며 해시 테이블은 key와 Value를 갖는 자료구조입니다. 주로 효율적인 검색에 활용되는데 이러한 해싱에서 발생할 수 있는 문제에는 어떤 것이 있을까요?

: 해시 함수는 해시 테이블의 키 값으로 레코드가 저장되어 있는 주소 혹은 색인을 산출하는 함수입니다. 순차검색에 비해 해시 테이블을 이용한 검색은 속도 측면에서 획기적이라고 할 수 있습니다. 그런데 데이터를 키로 간소화하여 저장한다는 아이디어는 좋지만 다른 내용의 데이터가 같은 키를 갖는다면 문제가 생길 것입니다. 이러한 상황을 해시 충돌이라고 하는데 같은 키값을 갖는 데이터가 생긴다는 것은 특정 키의 버킷에 데이터가 집중된다는 뜻이기도 합니다. 그래서 너무 많은 해시 충돌은 해시 테이블의 성능을 떨어뜨립니다. 그러므로 해시 함수를 잘 정의하여 해시 충돌을 최소화 하는 것이 성능 개선에 도움이 됩니다.

해시 테이블의 충돌 해결

해시 함수를 통해 만들어진 해시 주소가 중복되면 데이터 값이 충돌한다.

충돌이 발생할 경우 대처 방법이 필요하며 충돌 해결하는 방법은 해시 테이블에서 가장 중요한 문제이다.

대표적인 충돌 해결 방법은 체이닝과 개방 주소법이 있다.

체이닝

같은 주소로 해싱되는 원소를 모두 하나의 연결 리스트로 관리하다.

추가적인 연결 리스트 필요하다.

연결 리스트만 사용하면 되므로 복잡한 계산식을 사용할 필요가 개방 주소법에 비해 적다.

개방 주소 방법

충돌이 일어나더라도 어떻게든 주어진 테이블 공간에서 해결하여 추가적인 공간이 필요하지 않다.

개방 주소 방법은 해시 충돌이 일어나면 다른 버킷에 데이터를 삽입하는 방식이다.

빈자리가 생길 때까지 해시값을 계속 만들어낸다.

삽입, 삭제 시 오버헤드가 적고 저장할 데이터가 적을 때 더 유리하다.