



# Set 인터페이스의 해시기반 자료구조 활용 (1)

- MyHashSet을 통한  
실 세계 데이터 분석

# 학습내용

## 1 Set 인터페이스의 해시기반 자료구조를 이용한 실 세계 데이터 셋 분석

# 학습목표

- Set 인터페이스의 해시기반 자료구조를 활용하여 실 세계 데이터 셋 분석을 할 수 있다.
- 직접 구현한 자료구조와 다른 자료구조의 비교 분석 및 한계점 분석을 할 수 있다.

## 분석1

### 이메일 데이터에 존재하는 모든 ID 중복없이 유지하기

#### 다음 방법의 효율성 측정

모든 ID를 MyHashSet<Integer> idSet에 삽입  
(자동 중복 제거)

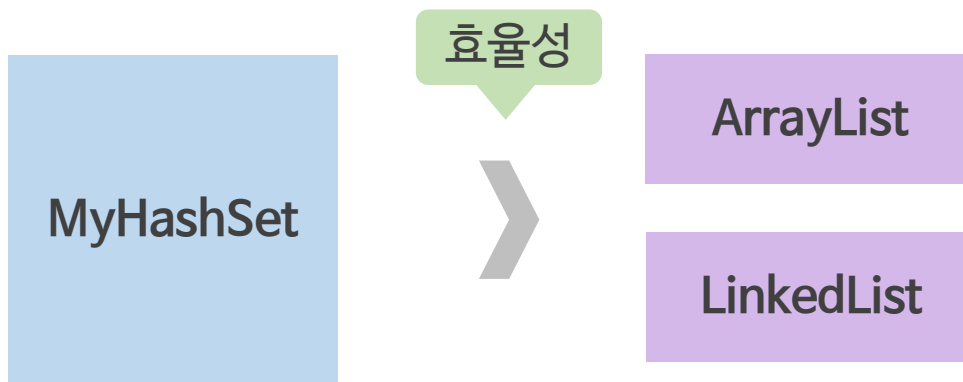
모든 ID를 HashSet<Integer> idSet에 삽입  
(자동 중복 제거)

모든 ID에 대해 ArrayList<Integer>에  
!contains(ID)라면 삽입

모든 ID에 대해 LinkedList<Integer>에  
!contains(ID)라면 삽입

## 분석1

이메일 데이터에 존재하는 모든 ID 중복없이 유지하기



## [분석 1]

**이메일 데이터에 존재하는  
모든 ID 중복없이 유지하기**

## 분석2

이메일 데이터에 존재하는 각 ID에 대해, 보낸 사람에 존재하는지 확인하기

모든 ID를 갖는 HashSet<Integer> idSet 준비

보낸사람 ID를 갖는 MyHashSet<Integer>  
senderSet2 준비

보낸사람 ID를 갖는 HashSet<Integer>  
senderSet 준비

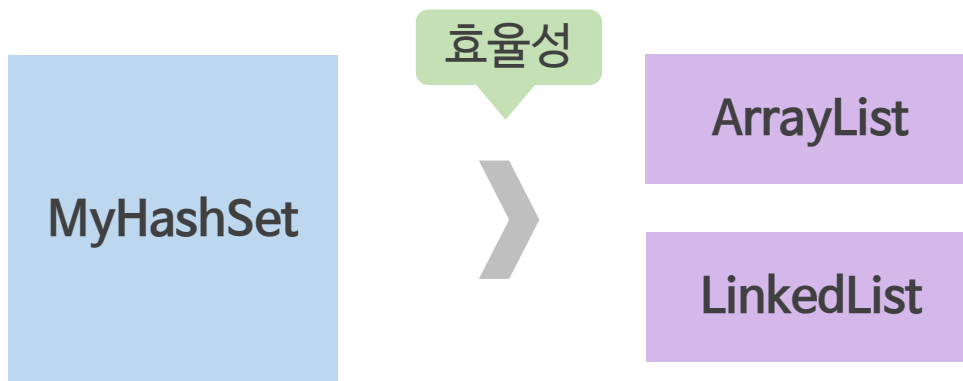
보낸사람 ID를 (중복없이) 갖는 ArrayList<Integer>  
senderArrayList 준비

보낸사람 ID를 (중복없이) 갖는 LinkedList<Integer>  
senderLinkedList 준비

idSet을 순회하며, 각 id에 대해 속도 측정

## 분석2

이메일 데이터에 존재하는 각 ID에 대해, 보낸 사람에 존재하는지 확인하기



## [분석 2]

**이메일 데이터에 존재하는  
각 ID에 대해, 보낸 사람에  
존재하는지 확인하기**



## 분석3

### 집합 관련 문제

이메일 네트워크에 참여한 사람의 수  
(합집합)

이메일을 보내기도 하고 받기도 한 사람의 수  
(교집합)

이메일을 보내기만 한 사람의 수  
(차집합)

[분석 3]

**집합 관련 문제**

# Remind

## MyHashSet을 이용하여 실 세계 데이터 셋 분석

이메일 데이터에 존재하는  
모든 ID 중복없이 유지하기

이메일 데이터에 존재하는  
각 ID에 대해, 보낸 사람에  
존재하는지 확인하기

집합 관련 문제