자료구조 프로젝트 보고서

2012210044 컴퓨터학과 이승훈

**1. 내가 선택한 자료구조**

내가 선택한 자료구조는 바이너리 서치 트리이다. 원래 레드블랙 트리를 이용하려 했으나 레드블랙트리를 구현할 자신이 없어서 조금 쉬운 바이너리 서치 트리를 사용하였다.

**2. 예상 성능**

우선 바이너리 서치 트리를 이용했기 때문에 search를 하는데 평균 O(nlogn)의 time complexity가 걸리지만 데이터를 다시 봤더니 user.txt의 경우 userID가 증가하는 방식으로 정렬되어 있었다. 즉 바이너리 서치 트리이긴 하지만 계속 right child로만 insert되기 때문에 balance가 없는, 즉 worst case라 볼 수 있다. 따라서 O(n^2)의 time complexity가 걸릴 것으로 예상된다.

하지만 사용자들의 tweetList는 worst case가 아니기 때문에 O(nlogn)의 time complexity를 가질 것이다.

**3. 성능 향상을 하려면**

우선 위에서 말했듯이 레드블랙 트리를 이용하면 서치를 할 때 O(nlogn)의 시간을 가지게 된다. 즉 자료구조를 바꾸게 되면 더 좋은 성능을 가지게 할 수 있고, 알고리즘 측면에서도 중복되는 검사는 하지 않도록 잘 조정하면 성능이 향상될 것이다.

**●Self Evaluation**

-submit a github account : 10  
-commit source code displaying menu : 10  
-commit the first draft of manual : 10  
-read data files : 20  
-statistics : 20  
-top 5 most tweeted words : 0  
-top 5 most tweeted users : 0  
-find all users who mentioned a word : 0  
-find all users who are friend of the above user : 0  
-top 5 strongly connected components : 0  
-find a shortest path from a user : 0