What data structure you chose and why

먼저 각 유저를 graph의 vertex 들로 나타낸 뒤 idN, screenname, tweetedword 등을 속성으로 가지도록 했습니다.

그 후 friend 관계를 읽어오면서 각 vertex들을 linked list로 연결하여 directed graph로 표현했습니다.

이는 탐색의 편리성과 친구관계 표현의 편의를 위한 것입니다.

또 uservertex 들을 mergesort를 이용해서 idN 순으로 정렬해 두었습니다.

What is your expected performance

0.데이터파일 읽어오기

1.통계표시

3.가장 많이 트윗한 유저 top 5 찾기

4.어떤 단어를 트윗한 모든 유저 찾기

5.3번 혹은 4번을 수행 한 후 그들의 친구 모두 찾기

6.어떤 단어의 멘션들 다 지우기

7.어떤 단어를 멘션한 유저 모두 삭제

는 모두 수행 가능하고

2. 가장 많이 트윗된 단어 top5

8. scc찾기

9. shortest path 찾기

는 수행 불가능합니다.

How would you improve the system in the future

시간을 입력받았으니 시간대별 통계라던지, 유저의 시간대별 사용 빈도 등을 분석할 수도 있고 샘플 데이터에 보니 욕설이 섞인 트윗이 있던데 욕설을 일정 횟수 이상 사용하는 유저들을 삭제하는 기능도 추가할 수 있을 것 같습니다.