자료구조 프로젝트

2009120090 경영학과 이병규

# cUser 클래스

유저의 정보를 담기 위한 클래스로, 문자열 변수 ID, 문자열 변수 screenName를 가지고 있다.

또한 스택형 자료구조인 std::vector를 사용하여 트위터 정보를 담는 vTweet과, 친구 정보를 담는 vFriend 변수를 가지고 있다.

# cTweet 클래스

트윗한 정보를 담는 클래스로, 트윗한 날짜를 담는 문자열 변수 date와, 하나의 트윗에 포함된 단어들을 담는 벡터를 가지고 있다.

# cWord 클래스

각 단어에 대한 정보를 가지고 있는 클래스이다. 어떤 유저가 해당 단어를 사용했는지를 알기 위해, 단어를 사용한 유저의 ID를 키 값으로 하여 자료를 관리하는 std::set으로 저장한다.

# cUserList 클래스

유저의 ID를 키값으로 갖는 해쉬형 자료구조인 map을 사용하여 유저의 정보를 저장하는 mapUser 변수를 사용하였다. 이는 고유한 정보인 ID를 바탕으로 보다 빠르게 값에 접근하기 위함이다. 마찬가지로 모든 단어에 대하여 단어를 키 값으로 하는 mapWord 변수를 사용하였다. 그 밖에 친구 관계의 수를 담는 정수형 변수와, 트윗 개수를 담는 정수형 변수를 가지고 있다.

# 초기 예상한 성능

사용자의 ID나, word를 키 값으로 하는 트리 구조를 계획하였다. 이진 탐색 트리 형태로 자료를 저장한다면 최대값과 최소값, 또는 단어의 발생 빈도를 측정하기 용이할 것이기 때문이다. 그러나 이진 탐색 트리를 구현하는 과정에 있어 애로사항이 발생하여, 지금과 같이 벡터와 set을 사용하게 되었다. 이로 인해 트리의 O(logN) 보다 탐색 성능이 저하되었다.

# 개선 방향

user와 word 정보를 트리화한다면, 최대값, 최소값 등을 찾는 데에 있어 월등한 성능을 낼 수 있을 것이다.