자료구조 프로젝트 리포트

2011210070 김정현

이번 자료구조 프로젝트를 진행하기 위해서 제가 선택한 자료구조는 Red Black Tree 이다. Red Black Tree는 BINARY SEARCH TREE에 비해 균형 잡혀있고 INSERT에 O(lg n)

DELETE에 O(lg n), SEARCH에 O(lg n)의 time complexity를 갖기 때문에 이번 프로젝트에 적절한 자료구조라고 할 수 있다. 텍스트 파일로부터 스트링을 읽어들여서 이 데이터들을 RED BLACK TREE에 저장한다. N개의 데이터를 저장하는 것은 time complexity가 O(nlogn)이다. Top5 most tweeted words와 Top5 most tweeted users를 찾는 것은 RED BLACK TREE에서 MAXIMUM과 PREDECESSOR를 이용하면 찾을 수 있는데 MAXIMUM과 PREDECESSOR 모두 O(logn)의 time complexity를 갖는다.

단어를 언급한 유저를 찾는 것과 이 단어를 언급한 유저의 친구를 찾는 것 모두 search이므로 O(nlogn)의 time complexity를 갖는다. 특정 단어를 찾아 지우는 것과 특정 단어를 언급한 유저를 지우는 것은 RED BLACK TREE에서의 DELETE이므로 O(logn)의 time complexity를 갖는다.

이번 과제는 word만을 다뤘는데 문장으로 된 실제 트위터의 내용도 다룰 수 있는 시스템으로 발전시켜 나가도록 할 것이다.

그리고 각 유저별로 쓴 단어를 분석해서 이 사람의 관심사가 무엇인지 사고 싶어하는 물건이 무엇인지를 알아볼 수 있는 시스템으로 발전시켜 나가는 것도 의미 있을 것 같다.