Hacettpe University
Department of Computer Engineering
Software Laboratory II
Assignment 2Report

Onur Tolga KESEMEN
21228511

April 7, 2017

Abstract

Subject: Complete Matching Words via Search Key
Submission Date: 24.3.2017
Due Date: 7.4.2017 (23:59:59)

1 Overview Of The Problem

In this assignment, we were expected to write a program that implements matching words via search keys automatically for a given set N term. A term is a query string and associated nonnegative weight.

The program will automatically completes the matching words via search key by sorting the terms by query string; binary searching to find all query strings that start with a given prefix; sorting the matching terms by weight; and prints out the top k matching terms in descending order of weight.(k value is given to program as argument.)

2 Solution

I have taken input text, k value and search key from commandline as arguments. Program include two classes; MatchingWords, Text. MatchingWords class contains main, sorting, binarySearch and other methods. Text class has two variable; weight, string and their getters-setters and compareTo methods.

In the beginning program reads input text then creates objects for per line(one pair of query string and nonnegative weight). Add them to ArrayList named list. Then sorts list with Heap Sort Algorithm with respect to string attribute of elements. Then using binary search to find matching words via search key and adding them to ArrayList query. Finally, sort the ArrayList query with respect to weight attribute of elements in descending order and prints them out.
Note: The program gives error if the nonnegative weight exceeds the range of type int in java. Because i declared them as integer not long. I couldn't do them as long.

2.1 Methods

sorting: Takes ArrayList(Text) list, int length, int precondition as parameter. Sorts the list with Heap Sort Algorithm and return list. Precondition is used for determining to list will be sort according to which attribute of elements.
precondition:0 – string precondition:1 – weight

binarySearch: Takes ArrayList(Text) list, String searchKey as parameter. Makes binary search in the list to find a matching term via searchKey. If it can find the term, adds it to ArrayList(Text) query. Then remove that term from list and doing binary search again until there is no match. Finally, return query.

reverse: Takes ArrayList(Text) list, int length as parameter. This method reversing the order of list. It is used because i need to reverse the order of the ArrayList(Text) query from ascending to descending.

2.2 Sample Input

Input format: *.txt (k value) (search key)
The program takes these inputs form commandline as arguments. Each file consists of an integer N followed by N pairs of query strings and nonnegative weights. There is one pair per line, with the weight and string separated by a tab. 'k value' is an integer and 'search key' is a string.

cities.txt
93827
14608512 Shanghai, China
13076300 Buenos Aires, Argentina
12691836 Mumbai, India
12294193 Mexico City, Distrito Federal, Mexico
11624219 Karachi, Pakistan
11174257 Istanbul, Turkey
10927986 Delhi, India
10444527 Manila, Philippines

wiktionary.txt
10000
5627187200 the
3395006400 of
2994418400 and
2595609600 to
1742063600 in
1176479700 i
1107331800 that
1007824500 was
879975500 his
...
392323 calves

2.3 Sample Output

cities.txt 3 Asb
74583 Asbest, Russia
30772 Asbe Teferi, Ethiopia
16116 Asbury Park, New Jersey, United States

alexa.txt 6 g
1000000 google.com
999987 google.co.in
999977 google.de
999974 google.co.uk
999973 google.fr
999969 google.co.jp

imdb-votes.txt 7 The
1282348 The Shawshank Redemption (1994)
1245679 The Dark Knight (2008)
939601 The Lord of the Rings: The Fellowship of the Ring (2001)
913272 The Lord of the Rings: The Return of the King (2003)
908680 The Matrix (1999)
885663 The Godfather (1972)
831456 The Dark Knight Rises (2012)

References

[1] Algorithms, 4th Edition, R. Sedgewick and K. Wayne, Addison-Wesley Professional, 2011