



INF1008
DATA STRUCTURES & ALGORITHMS
PART 1 ASSIGNMENT

Group 10

Name	Student ID
Wong Yok Hung	2202391
Christopher Kok	2203503
Chong Hou Wei	2201783
Vivian Ng	2203557
Methinee Ang	2202781

Contents

Question 1

Page 4

1.1 Problem Statement

4

Question 2

Page 5

2.1 Problem Statement

5

List of Figures

The Problem

US based phone numbers have been extracted from a bunch of documents and are available in a file, one phone number per line. Since people have many ways of writing phone numbers, the format of the numbers varies quite a bit. Numbers are always a minimum of 10 digits (3-digit area code and 7-digit number) but among the key variations, some numbers come with a “+1” prefix (country code for US is 1), other times there is only a “1” (as in “1800” numbers), the usual blocks are separated by either a space or a dash, the area code is sometimes within braces, sometimes all the digits are contiguous (no spaces), *etc.* Here are some key formats:

- +1-732-732-5555
- 1-732-732-5555
- (732) 732-5555
- 7327325555
- 732 7325555

For simplicity you can assume that the area code can be any 3 digit string that begins with a non-zero digit. You can also assume that the 7-digit number at the end may begin with a 0. So, for example the string “100 0000000” will be accepted as a phone number (for the purpose of this exercise!) Duplicates are allowed, in the same or different format.

Question 1

1.1 Problem Statement

Your first task is to find the median 10-digit phone number in a given input list (on the command-line). If there are an even number of phone-numbers, $2n$, then the n th and $(n + 1)$ th number from the sorted list must be output. For the purposes of this exercise, you should simply print out the median(s) as a 10-digit string, without any spaces. If there are 2 medians, they should simply be printed on the same line with comma separating them. Duplicates must be preserved. For example, in the list of 5 numbers above, the single median would be 7327325555 and so the output would be:

7327325555

If the input list were just the first 4 numbers, then there would be 2 medians, identical in this case, so the output would be:

7327325555,7327325555

Question 2

2.1 Problem Statement

2. Your second task is to extend the same program by allowing it to take TWO additional command-line inputs: a 10-digit number (that starts with a non-zero digit) and a K . If these additional inputs are provided, the program should print out the K unique numbers nearest (in terms of numerical difference) to the given 10-digit number. Duplicate numbers are considered ties and should also be printed. Thus, for example, if the program executable is called 'findnumbers' and the list of numbers above are in a file called 'phonescraped', then following is the command-line invocation for a number with $k = 1$: `Findnumbers phonescraped 7327325550 1` For this, the output would be:

7327325555

7327325555

7327325555

7327325555

7327325555

Since all numbers are tied and the closest to the input number.

Try to make the search as efficient as possible. As this will be part of the assessment.