THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

2019 SCHOLARSHIP EXAMINATION

PRACTICAL SECTION

DEPARTMENT	Computer Science
COURSE TITLE	Year 13 Scholarship
TIME ALLOWED	Six hours with a break for lunch at the discretion of the supervisor
NUMBER OF QUESTIONS IN PAPER	Three
NUMBER OF QUESTIONS TO BE ANSWERED	Three
GENERAL INSTRUCTIONS	Candidates are to answer ALL THREE questions. All questions are important. Answer as much of each question as you can. Plan your time to allow a good attempt at each question.
SPECIAL INSTRUCTIONS	Please hand in notes and answers to written questions, and a Pen Drive or DVD with your program/computer work for each question. In addition please make sure that a copy of each program stored as a plain text file. You cannot assume that the examiner has available any special software that might be required to read your files. Candidates may use any text or manual for reference during the examination. Candidates may not have access to the internet during the examination except for access to online documentation for the software they are using.
CALCULATORS PERMITTED	Yes

1. **Superannuation** (Spreadsheet Use)

In this question you are asked to use a spreadsheet to do calculations and to display the results. We expect that the spreadsheet will be used for all calculations unless the question states otherwise - you will be marked down for performing calculations by hand and directly entering the results. Your work will be graded on three criteria.

(i) The accuracy of your results.

(ii) The skill you show in making use of the capabilities of the spreadsheet.

(iii) The presentation of your results. We have deliberately not provided any instructions concerning layout or formatting. Your goal is to make the spreadsheet easy to use.

Spreadsheets are often used for financial 'simulation' – allowing people to create what-if situations and explore the consequences. For example a company might want to see what happened if petrol prices changed. They could make a spreadsheet with details of the costs of various parts of their operation, including estimates of costs that might be passed on by their suppliers, and try different petrol price values to see what happened. Simulations can be carried forward for several years to look at longer term effects of change. Governments too use simulations. In this question you are asked to take some information obtained from Statistics New Zealand and use it to explore an aspect of government expenditure.

National superannuation is a payment made by the government to support people in their old age. Payments begin on a person's 65th birthday. The amount paid depends on a person's circumstances, but for the purposes of this calculation we will use an average amount of $410 per person per fortnight. Payment is made until the person dies. In New Zealand, recent decades have seen a decrease in the birth rate, and an increase in life expectancy. The result is that fewer working taxpayers will be required to pay for the superannuation of an increasing population of aged people. The government is concerned about the viability of this situation and is considering changing the age at which payments start from the current 65 years to an age of 67 or 68. Your task is to provide a spreadsheet which can be used to explore some effects of such a change.

You have been provided with three files: PopulationByAge.csv, DeathRates.csv and BirthRates.csv. Check with your examination supervisor if those files are not available. Each of the files is formatted as 'comma separated value' (`.csv`). This means that they are simple text files with information values separated by commas. In each case the content and format of the file will be explained in the instructions following. The data in the files is adapted from information taken from the Statistics New Zealand web site.

The task is presented in stages. We recommend that you proceed stage by stage.

Stage A

The first file is 'PopulationByAge.csv'. Here are the first 12 lines of the file

```
Less than one year,58158
One year,58020
Two years,58719
Three years,59970
Four years,60054
Five years,64164
Six years,63531
Seven years,65085
Eight years,65289
Nine years,64563
10 years,64962
11 years,62151
```

The lines give the New Zealand population in 2018 by age. I.e. there were 58,158 babies less than one year old, 58,020 one year old children, etc.

Create a spreadsheet using the data from the file. Add a calculation to determine the total population of New Zealand in 2018.

Stage B

New Zealand National Superannuation is paid to all people of age 65 and over. The government is interested in the total cost of National Superannuation. Add a calculation to your spreadsheet to report the total cost.

The government is interested in varying the age at which superannuation payments start. Add a cell to your spreadsheet in which the starting age can be entered. Add to your spreadsheet so that it shows the new cost of superannuation for different starting ages. You should develop this calculation so that the current cost (starting at age 65) is still visible, to allow comparison.

Stage C

Sometimes simple costs don't tell the whole story. For example, if the population of New Zealand were doubled, but the fraction of the population at each age remained the same, the total cost of superannuation would double. That would probably not be a problem, because the number of taxpayers should also double and therefore the government should still be able to afford to pay for superannuation. On the other hand, if a new medical treatment were developed to double life expectancy after 65, the cost of superannuation would also double, but without an increase in the number of taxpayers. That could be difficult for the government.

Devise and add to your spreadsheet a way of reporting the cost of superannuation which would give the government a better indication of the 'affordability' of National Superannuation. Make sure that there is text on the spreadsheet briefly explaining your new cost report.

Stage D (Note: You may like to work on stages D and E at the same time)

The second file provided is 'DeathRates.csv'. It gives the likelihood of a person dying at different ages. The first few lines of the file follow. Each line has two values separated by commas. The first value is the start of an age range in years. The second value is the percentage chance of dying on or after that age, but before the age mentioned on the following line. In the data below, the first line gives the chance of dying before the age of 1. It is 0.3635% - a little over one chance in 300. The second line has the change of dying between the ages of 1 and 4 inclusive. The third is the chance of dying between the ages of 5 and 9 (inclusive). The last line in the file gives the chance of dying at or over the age of 110. You may assume that no-one lives beyond the age of 112 years.

```
0,0.3635
1,0.074800699
5,0.051748912
10,0.063338226
15,0.206741483
```

The 'DeathRates' file gives you part of the information needed to estimate how the population of New Zealand will change in years to come.

Extend your spreadsheet to report the cost of New Zealand superannuation each year until 2050, assuming that people die at the rates given, but that there are no births. This should lead to an aging and reducing population.

Stage E

The third file provided is BirthRates.csv. The whole file is as follows:

```
15 to 19,7.445
20 to 24,27.55
25 to 29,46.73
30 to 34,58.295
35 to 39,33.365
40 to 44,7.435
45+,0
```

This file gives the number of live births per 1000 people per year. The first line therefore states that there are 7.445 births each year for each 1000 people between the ages of 15 and 19 (inclusive).

Extend your spreadsheet to include births.

2. **Scores** (Careful and Accurate Programming)

Your programming work in this question will be assessed on two criteria:

(a) Completeness and accuracy of the program. It may be that this problem statement does not state exactly what the program should do under all circumstances. If you find a situation of that nature, choose a solution and write down, either on paper or in the comments of your program what the difficulty was and how you chose to resolve it.

(b) Good presentation. That is, it should make good use of programming language facilities, be well organised, neatly laid out, and lightly commented.

In this question you are asked to write a program to process results from the first rounds of matches in the Rugby World Cup. The Rugby World Cup is played in two phases. Your task is to analyse results from the first phase (the pool phase).

The following is an abbreviated version of the official rules of the Rugby World Cup.

The 20 Teams in the competition were allocated to four 'pools' of five Teams each. The pool phase consists of 'round robin's in which each team plays each other team in their pool once, giving a total of 10 matches played in each pool. Teams are awarded 'match points' for each match as follows:

- The team that wins is given 4 match points
- In the case of a draw, both teams are given 2 match points
- A team scoring 4 or more tries in a match gets 1 bonus match point (note that winning and losing teams in a match may both earn this bonus).
- The losing team is given 1 bonus match point if they lose the match by 7 points or less.

Cancelled Matches: Where a pool Match cannot be commenced on the day in which it is scheduled, it shall not be postponed to the following day, and shall be considered as cancelled. In such situations, the result shall be declared a draw and Teams will be allocated two Match points each and no score registered. For the avoidance of doubt, no bonus points will be awarded.

At the completion of the pool phase, the teams in a pool are ranked one through five based on their cumulative Match points, and identified respectively as winner, runner-up, third, fourth and fifth. If at the completion of the pool phase, two or more Teams are level on Match points, then the following criteria shall be used in the following order until one of the Teams can be determined as the higher ranked:

1. The winner of the Match in which the two tied Teams have played each other shall be the higher ranked.
2. The Team which has the best difference between points scored for and points scored against in all its pool Matches shall be the higher ranked.
3. The Team which has the best difference between tries scored for and tries scored against in all its pool Matches shall be the higher ranked.
4. The Team which has scored most points in all its pool Matches shall be the higher ranked.
5. The Team which has scored most tries in all its pool Matches shall be the higher ranked.
6. Should the tie be unresolved at the conclusion of steps 1 through 5, the rankings as per the updated Official World Rugby World Rankings will determine the higher ranked team.

The twenty teams qualifying for the Rugby World Cup and their world rankings are as follows:

The 20 qualifying teams in alphabetic order and their world rankings. The ID numbers have been added for you to use as program input, if you find them helpful.

ID	Team Name	World Ranking
1	Argentina	10
2	Australia	6
3	Canada	22
4	England	3
5	Fiji	11
6	France	8
7	Georgia	14
8	Ireland	4
9	Italy	12
10	Japan	7
11	Namibia	23
12	New Zealand	1
13	Russia	20
14	Samoa	15
15	Scotland	9
16	South Africa	5
17	Tonga	13
18	Uruguay	18
19	USA	17
20	Wales	2

Your program should analyse the results from one group in the World Cup. It should prompt for input and display results. A sample interaction with your program might look as shown on the next page (user input is underlined). Note however, that you are free to program the interaction in any way that solves the problem.

```
Rugby World Cup Scoring
Enter the team numbers of the teams in the group
:  5 4 8 10 13
In the match between Fiji and England
  how many points did Fiji score: 45
  how many tries did Fiji score: 6
  how many points did England score: 46
  how many tries did England score: 5
In the match between Fiji and Ireland
  how many points did Fiji score: 36
  how many tries did Fiji score: 4
  how many points did Ireland score: 50
  how many tries did Ireland score: 6
In the match between Fiji and Japan
  how many points did Fiji score: 12
  how many tries did Fiji score: 2
  how many points did Japan score: 33
  how many tries did Japan score: 3
In the match between Fiji and Russia
  how many points did Fiji score: 60
  how many tries did Fiji score: 10
  how many points did Russia score: 12
  how many tries did Russia score: 1
In the match between England and Ireland
  how many points did England score: 25
  how many tries did England score: 2
  how many points did Ireland score: 27
  how many tries did Ireland score: 3
In the match between England and Japan
  how many points did England score: 28
  how many tries did England score: 2
  how many points did Japan score: 34
  how many tries did Japan score: 6
In the match between England and Russia
  how many points did England score: 41
  how many tries did England score: 5
  how many points did Russia score: 20
  how many tries did Russia score: 1
In the match between Ireland and Japan
  how many points did Ireland score: 24
  how many tries did Ireland score: 2
  how many points did Japan score: 20
  how many tries did Japan score: 3
In the match between Ireland and Russia
  how many points did Ireland score: 38
  how many tries did Ireland score: 2
  how many points did Russia score: 12
  how many tries did Russia score: 1
In the match between Japan and Russia
  how many points did Japan score: 51
  how many tries did Japan score: 4
  how many points did Russia score: 22
  how many tries did Russia score: 2
All 10 matches were played (no cancellations)
The group winner was Ireland with 17 match points
```

3. **Language** (Problem Solving and Programming)

Your programming work in this question will be assessed on two criteria:

(a) Your approach to the problem. We will be looking at your work for evidence that you found good ways of storing the necessary data, and devised algorithms for finding and displaying the requested results. ***Please hand in any notes and diagrams which describe what you are attempting to program, even if you don't have time to code or complete it. You may include comments in your program, or write a description of your program to hand in.***

(b) The extent to which your program works and correctly solves the problem.

You may find that the programming language you use makes it difficult to produce output as shown in the example implementation steps below. If this is the case, feel free to build your program in a way that suits your circumstances.

Development of natural language processing (NLP) software – programs that process and/or understand text written in a human language – is a major area of computer science research. Perhaps the most successful NLP software yet created is 'Watson', produced by IBM to process English text and answer questions. It successfully competed against human players to win the quiz game Jeopardy, and has been applied to the interpretation of medical texts and papers. Other applications of NLP include indexing of web pages, and automatic translation from one language to another. One of the research problems in NLP is language identification – deciding in which language a text is written.

Your task in this question is to build a program that reads lines of text written in either Māori or English, and for each line, decides which of those languages is being used. Your goal is to produce an efficient algorithm that could be used to process many lines of text quickly. For that reason your program will not attempt to understand the texts. Instead it will use information about the languages to make rapid decisions. Those decisions will not always be right. It is up to you to test and evaluate two different approaches. The stages of the question will guide you through this work.

You have been provided with four files: mlines.txt, elines.txt, m.txt and e.txt. Check with your examination supervisor if those files are not available. mlines.txt and elines.txt hold lines of text in Māori and English respectively. They are taken from the Māori and English versions of the Treaty of Waitangi respectively and are to be used to test your algorithms. m.txt and e.txt hold 500 of the most commonly used words of Māori and English respectively.

Stage A: Finding words in a line

Your program is to read and identify the language used in single lines of text. You have been provided with sample files with lines in Māori and English (mlines.txt and elines.txt).

Begin by writing a program to find the words in a line of text. You may read the line from a file, or type it in yourself. Your program should break the line into separate words and display the words. You will need to think about the way you identify words. What characters can occur in a word? What characters occur between words?

```
Enter a line of text
>> Sample line.  It has six words.
Word 1: Sample
Word 2: line
Word 3: It
Word 4: has
Word 6: six
Word 7: words
```

Stage B: Reading a file

Extend your program to read all the lines in a file. You can test your program on mlines.txt and elines.txt. For example, reading mlines.txt could give output like this

```
Line 1
>> KO WIKITORIA te Kuini o Ingarani i tana mahara atawai
Word 1: KO
Word 2: WIKITORIA
Word 3: te
Word 4: Kuini
Word 5: o
Word 6: Ingarani
Word 7: i
Word 8: tana
Word 9: mahara
Word 10: atawai

Line 2
>> ki nga Rangatira me nga Hapu o Nu Tirani i tana hiahia
Word 1: ki
Word 2: nga
. . .
```

Stage C: First algorithm

Remember the goal is to decide which language is used in each line of text. The sample files you have been given are pure Māori and pure English. You can create your own test files in which some lines are in Māori and some are in English.

The first idea for building an algorithm to decide on the language of a line is based on the observation that written Māori does not use all of the letters in the English alphabet. In

fact it only uses the letters a, e, h, i, k, m, n, g, o, p, r, t, u, and w. Therefore, lines that have words using other letters are likely to be in English.

Extend your program to check the letters used in the words on each line. Where possible, for each word, determine the language. Note: letters in use may be upper case or lower case.

Test your program. Write a paragraph reporting on the success of the algorithm. Your paragraph should explain the circumstances in which this algorithm might be expected to work, and the circumstances in which it might be expected to fail.

Stage D: Second algorithm

The second idea for building an algorithm is based on the fact that languages have very commonly used words. For example 'the' is the most commonly used word in English text. The idea for the algorithm is to use a list of the most commonly used words in each language and count uses of those words. The files m.txt and e.txt have the 500 most commonly used words in Māori and English respectively. The words in each file are sorted into frequency order – the most used word first. Therefore, if you wanted to use the most commonly used 100 words, you could just take the first 100 words from each file.

Extend your program to check words used in each line against lists of commonly used words. You will need to decide how to use the information gained to decide on the language. Again you will need to take into account upper and lower case letters.

Test your program. Write a paragraph reporting on the success of this algorithm. Your paragraph should explain the circumstances in which this algorithm might be expected to work, and the circumstances in which it might be expected to fail.

Stage E: Comparison

Write a paragraph comparing the algorithms. Which to you consider to be the best?

Notes and Acknowledgements

- Currently accepted best practice writing in the Māori language is to use macrons – the bar over the 'a' in the word Māori – to better represent the sounds of the words. In this question all text in the sample files has had macrons removed to simplify the programming.
- The common words lists have been reproduced from http://tereomaori.tki.org.nz/Teacher-tools/Te-Whakaipurangi-Rauemi/High-frequency-word-lists and https://world-english.org/english500.htm.
- The text of the treaty of Waitangi is taken from https://nzhistory.govt.nz/politics/treaty/read-the-treaty/maori-text