Research Fellow selection process

Software development exercises

We would you to carry out the following exercises as part of the selection process:

1. Coding exercise – writing a piece of code to fulfil a task.

2. Presentation – a 5 minute presentation to be give during your interview.

## 1. Coding exercise

We would like you to write a piece of code in Python, R or a similar programming language to fulfil the following task. Please ensure that the logic of your code is clear, including comments as appropriate.

Please write a function to extract specific medication codes and associated medication dosages from a text document. Your function should accept a ***string*** as input, and should output a ***list*** of medication items, where each item is a sublist consisting of (1) the medication code, as an ***integer***, and (2) the dosage, as a ***string***.

Medication names should be converted into an integer code, according to the following table:

|  |  |
| --- | --- |
| **Medication name** | **Code (integer)** |
| paracetamol | 1 |
| panadol | 1 |
| aspirin | 2 |
| penicillin | 3 |

The text document will be provided as a single string with line breaks denoted by the newline character (“\n”). It will contain any number of medications with dosages, each on a separate line. The medication name is at the beginning of the line, followed by one or more spaces, then the dosage instructions in the form of free text. Medications are not case sensitive. There may also be medications that are not in the table above, which should be ignored.

The medication section is ended by a blank line or a line which contains only spaces, or the end of the document. Anything below a line which contains only spaces should be ignored.

Example text:

Aspirin 75mg daily

PANADOL 1000mg WHEN NEEDED MAX FOUR TIMES DAILY

Atorvastatin 20mg daily

Penicillin 500mg daily

Paracetamol

History:

Feeling well up until yesterday, then he suddenly developed a severe headache.

Example output (Python):

[[2, '75mg daily'], [1, '1000mg WHEN NEEDED MAX FOUR TIMES DAILY'], [3, '500mg daily'], [1, '']]

Please submit your script before the interview.

## 2. Presentation

Please prepare a presentation of up to 5 minutes to explain the code and algorithm that you have written for exercise 1, in a way that is understandable for people with a general understanding of computer science but who may not be proficient in the programming language you have used.

Describe how you would further develop the program to:

* extract the names of a larger range of drugs
* handle mis-spelt names of drugs

You will be asked to give this presentation at the start of your interview. You may use slides if you wish.