

The Puzzle

Before a certain txt file can be analysed it needs to be cleaned of 3 text strings which have unexpectedly corrupted the txt file.

The 3 corrupt strings are:
"123456789",
"corruptcorrupt" and "Where is Wally?"

Your task is to write a program to clean the txt file of the 3 corrupt strings.

```
web design corruptcorrupt
File systems and relational
databases
Design process
3.1 understand and list user needs/requirements before def:
3.2 create a basic relational database to store and retriev
3.3 use appropriate programming languages to develop an in
of users' needs
APPLIED LEARNING TASK 2: ANALYTICS
Hypothesising, making predictions, examining evidence, rec
In this applied learning task, students will identify an i
highlight the salient information and inform future decisio
deepen their understanding of the practices and principles
Students learn about:
Students should be able to:
Analytics Abstraction
Data collection and analysis
Interpretation of data Where is Wally?
```

Abstraction

We want to develop an algorithm that can :

1. Read in the file contents as a string or a list or some structured data type
2. Recognise each of the 3 corrupt strings
3. Remove a corrupt string

Writing your Thinking

Take 5 minutes to think about how you tackled this problem.

- Did you break it down to a smaller problem such as replacing one particular corrupt string first?
- **Did you look for patterns in order to make it easier to solve?**
- Did you try different scenarios to verify your solution worked by inserting extra corrupt strings in the txt file?

Using Think-Pair-Share-Square (TPSS), go through how you and your partner were thinking about how to solve the problem.

Pseudo-Code

#Initialisation

create a list of corrupt_strings;

#read the corrupted txt file

read in the file_contents from the txt file;

#iterate through the list of corrupt_strings

```
for each word in the list of corrupt_strings {  
    remove that word in the file_contents  
}
```

#write out the clean contents to a new file

write out the edited file_contents to the new txt file;

Extending your program

Can you split the file contents into a list of individual strings or words?

Once this is done, each word can be analysed for its length or frequency of occurrence, hence opening up the possibility of some data analytics. There is code to do this in the repl.it sandbox below. Just click on the repl.it symbol to view more.

LO 3.5 students should be able to structure and transform raw data to prepare it for analysis