

## **Lab8 – Sets and Dictionaries**

### **Lab Objectives**

The aim of this worksheet is to use the data structures sets or maps as ways of storing and accessing data. It is also an opportunity to reflect of previous data structures we have learnt such as lists or arrays.

### **Question 1**

Write a function that takes as input a text file and returns a tuple formed by a histogram - a set of frequencies for all the words in the text file - and the total number of words in the file. Assume the text file does not contain punctuations.

### **Question 2**

A police officer has received few reports. But they want to pay attention to reports including a specific word.

1. Write a function `word_search`, which takes as inputs a set of reports and a keyword to return all the reports that contain the given keyword. Your function should meet the following criteria
  - Do not include documents where the keyword string shows up only as a part of a larger word. For example, if she were looking for the keyword "closed", you would not include the string "enclosed."
  - She does not want you to distinguish upper case from lower case letters. So the phrase "Closed the case." would be included when the keyword is "closed"
  - The punctuation was removed from the text so no commas, full stop, etc.

**Example:** If you have

`docs = {"doc1.txt", "doc2.txt", "doc3.txt"}` wherein

- doc1 contains "The Learn Python Challenge Casino",
- doc2 contains "They bought a car",
- and doc 3 contains "Casinoville", then, your function

`word_search(docs, 'casino')` would return  
`{'doc1.txt'}`

2. As we would like to also enable multiple searches, write a function, `multi_word_search`, which takes as inputs a set of documents and a list of keywords to return a dictionary mapping each keyword to a list of documents that contain it.

### **Question 3**

A game player is trying to perform an analysis to determine the best items for winning races on the Mario Kart circuit. He has some data in the form of list of dictionaries that look like:

```
[
{'name': 'Peach', 'items': ['green shell', 'banana', 'green
shell',], 'finish': 1},
{'name': 'Bowser', 'items': ['green shell',], 'finish': 3},
# Sometimes the racer's name wasn't recorded
{'name': None, 'items': ['mushroom',], 'finish': 1},
{'name': 'Toad', 'items': ['green shell', 'mushroom'], 'finish':
2}
]
```

- “items” is a list of all the power-up items the racer picked up in that race, and
- “finish” was their placement in the race (1 for first place, 3 for third, etc.).

Write a function `winners` that takes in a list of dictionaries as given above and return a dictionary mapping items to the number of times those items were picked up by racers who finished in first place. If the racer’s name is missing, replace it by ‘Unknown’ in the dataset.