

Practical Work 4: Word Count

Vu Dinh Anh

February 2021

Word Count

Files: `vc.py`, `poem.txt`

Usage: `python vc.py poem.txt`

Expected output:

```
[('And', 1),  
 ('Roses', 1),  
 ('Sugar', 1),  
 ('Violets', 1),  
 ('are', 3),  
 ('blue', 1),  
 ('is', 1),  
 ('red', 1),  
 ('so', 1),  
 ('sweet', 1),  
 ('you', 1)]
```

Mapping

```
def _map(file):  
    word_map = []  
    with open(file, 'r') as f:  
        for line in f:  
            words = line.strip().split()  
            word_map += [(word, 1) for word in words]  
    return word_map
```

Sorting

```
def _sort(word_map):  
    word_map.sort(key = lambda word: word[0])  
    return word_map  
# A little bit of cheating here
```

Reducing

```
def _reduce(word_map):
    word = None
    cur_word = None
    cur_count = 0
    reduced_word_map = []

    for word, count in word_map:
        if cur_word == word:
            cur_count += count
        else:
            if cur_word is not None:
                reduced_word_map += [(cur_word, cur_count)]
            cur_word = word
            cur_count = count

    if cur_word == word:
        reduced_word_map += [(cur_word, cur_count)]

    return reduced_word_map
```