

University of Lincoln
School of Computer Science
CMP2801M – Advanced Programming
Workshop 2
Lecturer: Dr Ayse Kucukyilmaz

In this homework, you will implement a program that represents the summary of a given document. In particular, the summary will include the number of sentences, the number of words and the frequency of each word in the document. For this purpose, you will implement two classes: *DocSummary* and *WordFreqPair*. Details of each class are explained below.

There should be four files: *WordFreqPair.h*, *WordFreqPair.cpp*, *DocSummary.h* and *DocSummary.cpp*. You should put your class definitions and their implementation in the correct files. In order to run your program, you should use the given file which has the file name “main.cpp”.

WordCountPair Class

WordFreqPair class holds a word and its count. This class will be used in the *DocSummary* class to store the words and its counts in a given document.

Member Variables	Access modifier	Description
word	private	is a string that holds a word
count	private	is an integer that holds the number of occurrences of the word in the document

Member Functions	Access modifier	Input parameters	Return value	Description
Constructor	public	void	-	Defaults constructor. Sets word to null string and count to zero.
Constructor	public	std::string, int	-	Takes a word and its count as parameters and initialize the word and count member variables. Sets count to one if the second parameter is not supplied.
setWord	public	std::string	void	Sets word variable with the input
getWord	public	void	int	Returns word
setCount	public	int	void	Sets count variable with the input

getCount	public	std::string	void	Set count variable with the input
----------	--------	-------------	------	-----------------------------------

DocSummary Class

Member Variables	Access modifier	Description
numberOfWords	private	is an integer that indicates the number of words in the document
numberOfSentences	private	is an integer that indicates the number of sentences in the document
filename	private	is a string that shows the name of the document
wordList	private	is a vector of WordFreqPair objects. Each entry of the vector holds a WordFreqPair object that keeps the word and its count.

Member Functions	Access modifier	Input parameters	Return value	Description
Constructor	public	std::string	-	Takes document name as input and initializes class variables
printSummary	public	void	void	Prints the number of sentences, the number of words and each word with its frequency (see format at the end of the assignment)
increaseSentenceCount	private	std::string	void	Takes a word as an input and increases the variable <i>numberOfSentences</i> if the given word contains a ".", "!" or "?" symbol.
addWord	private	std::string	void	Adds a given word into the wordList vector if it does not exist in the vector. Otherwise it increases the corresponding count entry in vector.
analyseDocument	public	void	void	Reads the document word by word and calls addWord, removePunc, updateSentenceCount functions in the correct order to fill in the object properties

You are provided with an example input file and main.cpp, which should be your driver programme. Your output should have the exact same format given below. Here, the frequency of a word can be computed as the number of occurrences of the word in the file divided by the total number of words in the file.

```
number of sentences is 3
number of words is 40
Today    1
is       2
Sunday   1
For      1
the      4
first    2
time     2
they     1
took     1
me       1
out      1
into     1
sun      1
today    1
And      1
for      1
in       1
my       1
life     1
I        1
was      1
aghast   1
that     1
sky      1
so       3
far      1
away     1
and      2
blue     1
vast     2
```