CS101 Additional Homework

Misspelling Corrector

Due Date : Jun 22, 2015 (Until 23:59)

- No delayed submission allowed -

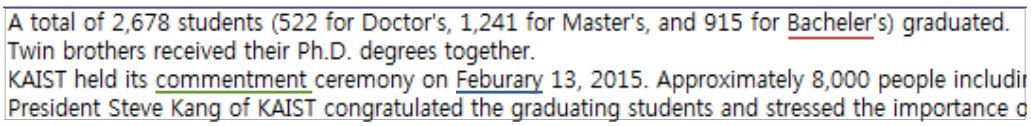
Please read the homework description carefully and make sure that your program meets all the requirements stated. The homework is an individual task. You can discuss the problem with your friends but you must not program together. **You will get F on the entire course if your homework includes any plagiarism**.

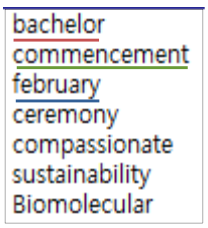
Goal

Implement 'Misspelling Corrector'.

In this homework, three text files (kaist1.txt, kaist2.txt, kaist3.txt) and one dictionary file (dic.txt) are given as below

In the text files, there are some (intentionally) misspelled words. In the dictionary file, there are correct words corresponding to the misspelled words.





The objective of 'Misspelling Corrector' is to produce corrected contents for given incorrect input using the given dictionary file. You must follow the next requirements for implementation.

Requirements

Your program needs to meet the following requirements:

(1) You must implement 3 classes - Word, Text, SpellingCorrector - at least. You may define your own classes to implement additional functions. In that case, you should describe details of the classes in your report.
(2) During implementation, you must keep the relations of 3 classes.
- 'Word' class represents an english word (string) with additional functions. Word object is an elementary user-defined datum in this homework.
- 'Text' class represents a collection of english words. This class has a list of Word objects as a member variable. The objective of Text class is to make Word objects from a text file.
- 'SpellingCorrector' class represents a misspelling corrector. This class has a Text object as a member variable. The objective of SpellingCorrector class is to produce a new misspelling-corrected Text object from an other Text object.

(3) You must implement 12 methods of 3 classes as follows:

Class	Method	Input	Output	Role
Word	__init__	word : An english word (string)	None	Initialization of a member variable 'data' with a parameter 'word'. The type of the variable must be string.
	__str__	None	A representative string of an object ex) Word : kaist 2014	Making human-readable description of an object.
	getData	None	An word data which is initialized by __init__()	Return the value of data variable.

Word	compareTo	other : An other Word object (Word)	The difference between two Word objects	Specifically, the difference between two Word objects is an **edit distance** of two english words.
Text	__init__	None	None	Initialization of a member variable 'content' with an empty list.
	__str__	None	A representative string of an object ex) File : kaist3.txt (172 words)	Making human-readable description of an object.
	load	textFilePath: A text file path which stores misspelled words	None	Load text from the file and save them into a list of Word objects. Assign the list as a member variable 'content'.
	getContent	None	A list of Word objects which is initialized by load()	Return the value of the member variable 'content'.
	setContent	words: A list of Word objects	None	Set the value of the member variable 'content' with the value of words.
	printContent	None	None	Print content of the Word object as human-readable format.
SpellingCorrector	__init__	dictionaryFile: A text file path which stores correct words	None	Initialization of a member variable 'correctedText' whose type is a Text class.
	correct	txt: A Text object	New Text object which has misspelling-corrected content	Correct misspelled words of given Text object and return new (corrected) Text object.

※ To implement 'compareTo' method of Word class, it is necessary to know about 'edit distance'. Visit the following link: http://en.wikipedia.org/wiki/Edit_distance

You can also refer to various implementations on the internet, but **beware of plagiarism**. It is recommended to explain about your implementation with comments and report.

※ In 'correct' method of SpellingCorrector class, your task is usually,
(1) Compare the content of a Text object with the content of the dictionary
(2) If an word in the text is very similar to an word in the dictionary, you can assume that the word is misspelled. Correct it.
(3) With corrected content, create new Text object and return it.

※ Hardcoding is not allowed. If you do hardcoding, you will get severe penalty.

Demonstration

We will supply a template file which includes a set of skeleton classes and main() method. The result of main() should be as below

- Code

```
# ------------------------------------------------------------
# Word object test code
# ------------------------------------------------------------

# Create an word object using a constructor
w1 = Word("kaist 2014")
w2 = Word("kiast 2015")

# Test getData() method
print w1.getData()
print w2.getData()

# Test __str__() method
print w1
print w2

# Test compareTo() method
print "Difference : " + str(w1.compareTo(w2))
```

- Expected Output

```
kaist 2014
kiast 2015
Word : kaist 2014
Word : kiast 2015
Difference : 3
```

- Code

```
# ------------------------------------------------------
# Text object test code
# ------------------------------------------------------

# Create a text object using a constructor
txt1 = Text()

# Test load() method
txt1.load("kaist1.txt")

# Test __str__() method
print "-------------------------------------------------"
print txt1
print "-------------------------------------------------"

# Test printContent() method
txt1.printContent()
print "-------------------------------------------------"
```

- Expected Output

```
-------------------------------------------------
File : kaist1.txt (116 words)
-------------------------------------------------
A total of 2,678 students (522 for Doctor's, 1,241 for Mast
Twin brothers received their Ph.D. degrees together.
KAIST held its commentment ceremony on Feburary 13, 2015. A
President Steve Kang of KAIST congratulated the graduating
-------------------------------------------------
```

- Code

```
# ------------------------------------------------------
# SpellingCorrect object test code
# ------------------------------------------------------

# Create a spelling corrector object using a constructor
corrector = SpellingCorrector("dic.txt")

# Test correct() method
txt1_corrected = corrector.correct(txt1)

print "-------------------------------------------------"
txt1_corrected.printContent()
print "-------------------------------------------------"
```

- Expected Output

```
-------------------------------------------------
A total of 2,678 students (522 for Doctor's, 1,241 for Maste
Twin brothers received their Ph.D. degrees together.
KAIST held its commencement ceremony on February 13, 2015. A
President Steve Kang of KAIST congratulated the graduating s
-------------------------------------------------
```

Evaluation

Your program will be tested by the TAs. They will check if your program implements the requirements properly. A program template, 'yourid.py' and the three data file (kaist1.txt, kaist2.txt, kaist3.txt) and one dictionary file (dic.txt) are provided. It is recommended to use the template file to implement the game program.

Submission

You need to submit your program source file which is named with your student ID (e.g., 20151234.py). You also need to submit the report about the implementation. The report file also needs to be named with your student ID (e.g., 20151234.doc or 20151234.pdf).

Please zip your homework files into “yourid.zip” (e.g., 20151234.zip), and submit it via the CS101 homepage.