# Homework 3

Upload your source code file from *DEUZEM SAKAI* until **December 18, 2022, 23:55.**

Upload only a single *.cs file.

The name of the file    :  **number_name_surname.cs**
For example            :  2022510028_mehmet_ali_yildirim.cs

## Question Answering System

Write a simple question answering program like [https://www.answers.com/](https://www.answers.com/). This problem falls under the domain of *Artificial Intelligence*.

The program will read the questions from a text file (*questions.txt*), find their answers by using a corpus (*corpus.txt*), and print the answers on the screen.

The questions and corpus may contain the following <u>punctuation marks</u> and <u>mathematical operators</u>:
`'.', ',', ';', '’' '"', '?', '+', '-', '*', '/'`

Make sure your program works well for uppercase and lowercase English letters.

The program will answer the questions according to the following three rules.

**Rule 1-** If the question starts with `"What is the result of expression"`, the program will find the result of mathematical expression and print it on the screen.

Assume that the expression will include a single type of operator (only + or only - or only * or only /).

**Rule 2-** If the user asks a question to the computer with the phrase `"What are the top10 words in the pattern"`, the program will find and print the words that corresponds to this *pattern*.

The symbol "-" corresponds to only one letter.

**Rule 3-** The program should compare the words in the question with the words in the text in the corpus, and find the most similar (probable) answer. The number of common words between question and text is the fundamental criteria. If two or more texts have the same maximum count, the program will print all of them.

To answer a question, at least two words should match, except stop words. If there is no any matching, print `"No answer"`

string[] stop_words = {"a", "after", "again, "all", "am", "and", "any", "are", "as", "at", "be", "been", "before", "between", "both", "but", "by", "can", "could", "for", "from", "had", "has", "he", "her", "here", "him", "in", "into", "I", "is", "it", "me", "my", "of", "on", "our", "she", "so", "such", "than", "that", "the", "then", "they", "this", "to", "until", "we", "was", "were", "with", "you"}

This homework will be graded by Res.Asst.Dr. Özge KART.

You can ask your questions to her from the "**FORUM → Homework 3 - Questions**" part of the *DEUZEM SAKAI* software.

## Sample Files

corpus.txt

```
Shakespeare is the author of Hamlet.
A soccer team has 11 players.
Soccer is the biggest participant sport in the world.
An anagram is a word or phrase that's formed by rearranging the letters of another word or phrase.
LCD television is brighter than LED television. Another big difference is the quality.
Jennifer Lopez is 53 years old.
Shakira is 45 years old.
Izmir is in Turkey.
Ankara is in Turkey.
Istanbul is in Turkey.
Before program, draw flowchart diagram.
The river Nile is 6695 kilometers long.
```

```
Nile is the longest river in the world.
Behind of the door, she has four poor boar sounding like your roar.
1024 KB is 1 Megabyte. 1024 MB is 1 Gigabyte. 1024 GB is 1 Terabyte.
Thomas Edison is an American inventor who developed electric light bulb.
A poem is a literary composition, written by a poet.
```

questions.txt

```
What is the biggest participant sport in the world?
What is the result of expression 2x3+5678x2+4x1234+21x71+5x2 for x=1?
Where is Izmir, Turkey or Italy?
What are the top10 words in the pattern ---g-am?
What is the biggest difference between an LED and LCD television?
What is the result of expression 2x3*12x2*4x3 for x=2?
How old is Jennifer Lopez?
What are the top10 words in the pattern ---?
Who is Thomas Edison?
What is the result of expression 3x3/12x2 for x=6?
How long is the river Nile?
What are the top10 words in the pattern -o-R?
What is your name?
Which sport has 11 players on a team?
```

**Sample Output:**

```
Answers

Soccer is the biggest participant sport in the world.
The result is 5710
Izmir is in Turkey.
program diagram anagram
LCD television is brighter than LED television. Another big difference is the quality.
The result is 24576
Jennifer Lopez is 53 years old.
the old has who she lcd led big
Thomas Edison is an American inventor who developed electric light bulb.
The result is 1.5
The river Nile is 6695 kilometers long.
door four poor boar your roar
No answer.
A soccer team has 11 players.
A basketball team consists of 11 players.
```

**Notes:**

**1.** Your program must work correctly under all conditions. Try to control all possible errors.

**2.** You should use meaningful variable names, appropriate comments, and good prompting messages.

**3.** If you are late, your grade will be decreased 10 points for each day.
   After five days, your assignment will not be accepted.

**4.** If you want, you may write your own "*procedures / functions*".

**5.** Assignment must be your <u>individual work</u>.
   **Cheating** is strictly prohibited.
   If any cheating occurs, your assignment will be graded with **zero (0)**.
   A software will be used to automatically detect the similarities between students' source-codes.