# Exercises: Text-Processing and Regular Expressions

Problems for exercises and homework for the ["Technology Fundamentals with C#" course @ SoftUni](https://softuni.bg/modules/57/tech-module-4-0).

You can check your solutions in [Judge](https://judge.softuni.bg/Contests/1217).

## Valid Usernames

Write a program that **reads user** names on a **single** line (joined by **", "**) and **prints** all **valid usernames**.

A valid username is:

* Has **length** between 3 and 16 characters
* **Contains** only letters, numbers, hyphens and underscores
* Has **no redundant symbols** before, after or in between

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| sh, too\_long\_username, !lleg@l ch@rs, jeffbutt | jeffbutt |
| Jeff, john45, ab, cd, peter-ivanov, @smith | Jeff  John45  peter-ivanov |

## Character Multiplier

Create a **method** that takes **two strings** as arguments and returns the **sum** of their **character codes,** **multiplied** (multiply **str1[0]** with **str2[0]** and add to the total sum). Then continue with the next two characters. If one of the strings is **longer** than the other, **add** the **remaining** character codes to the **total** **sum** without multiplication.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| Gosho Pesho | 53253 |
| 123 522 | 7647 |
| a aaaa | 9700 |

## Extract File

Write a program that reads the path to a file and **subtracts** the **file name** and its **extension**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| C:\Internal\training-internal\Template.pptx | File name: Template  File extension: pptx |
| C:\Projects\Data-Structures\LinkedList.cs | File name: LinkedList  File extension: cs |

## Caesar Cipher

Write a program that returns an **encrypted version** of the same text. Encrypt the text by **shifting** **each character** with **three** positions **forward**. For example: **A** would be replaced by **D**, **B** would become **E**, and so on. Print the **encrypted** **text**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| Programming is cool! | Surjudpplqj#lv#frro$ |
| One year has 365 days. | Rqh#|hdu#kdv#698#gd|v1 |

## Multiply Big Number

You are given **two lines** – the **first** one can be a really **big** number **(0 to 1050)**. The **second** one will be a **single** digit number **(0 to 9)**. You must display the product of these numbers.

Note: do not use the BigInteger class.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| 23  2 | 46 |
| 9999  9 | 89991 |
| 923847238931983192462832102  4 | 3695388955727932769851328408 |

## Replace Repeating Chars

Write a program that reads a string from the console and **replaces** any **sequence of the same letters** with a **single** **corresponding letter**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| aaaaabbbbbcdddeeeedssaa | abcdedsa |
| qqqwerqwecccwd | qwerqwecwd |

## String Explosion

Explosions are marked with **'**>**'**. Immediately after the mark, there will be an **integer**, which signifies the **strength** of the explosion.

You should **remove** x **characters** (where x is the **strength** of the explosion), **starting after** the punch **character** (**'**>**'**).

If you find **another** explosion mark (**'**>**'**) while you’re deleting characters, you should **add** the **strength** to your **previous** **explosion**.

When all characters are processed, **print** the string **without** the **deleted** **characters**.

You should **not** delete the **explosion** character – **'**>**'**, but you should **delete** the **integers**, which represent the **strength**.

You will receive a **single** **line** with the string. Print what is left from the string after the explosions.

### Constraints

* You will **always** receive a **strength** for the punches
* The path will consist only of letters from the **Latin** **alphabet**, **integers** and the char **'**>**'**
* The strength of the punches will be in the interval [0…9]

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comments** |
| abv>1>1>2>2asdasd | abv>>>>dasd | 1st explosion is at index **3** and it is with **strength** one. We delete **only** the **digit** **after** the explosion character. The string will look like this: abv>>1>2>2asdasd  2nd explosion is with strength **one** and the string transforms to this: abv>>>2>2asdasd  3rd explosion is with strength two. We delete the digit and we find **another** explosion. At this point the string looks like this: abv>>>>2asdasd.  4th explosion is with strength two. We have **1** strength **left** from the previous explosion, we **add** the strength of the **current** explosion to what is **left** and that adds up to a **total** strength of **3**. We **delete** the next **three** **characters** and we **receive** the **string** abv>>>>dasd  We do **not** have **any more explosions** and we print the result: abv>>>>dasd |
| pesho>2sis>1a>2akarate>4hexmaster | pesho>is>a>karate>master |  |

## Match Numbers

Write a program that finds all **integer** and **floating-point numbers** in a string.

A number has the following characteristics:

1. Has either a **whitespace** before it or the **start** of the string (match either **^** or what’s called a [positive lookbehind](http://www.regular-expressions.info/lookaround.html)). The entire syntax for the **beginning** of your **RegEx** might look something like “(^|(?<=\s))”.
2. The number might or might not be negative, so it might have a hyphen on its left side (“-“).
3. Consists of **one or more digits**.
4. Might or might not have **digits after the** **decimal point**
5. The decimal part, if it exists, consists of a period (“.”) and **one or more digits** after it. Use a **capturing group**.
6. Has either а **whitespace** before it or at the **end** of the string (match either **$** or what’s called a [positive lookahead](http://www.regular-expressions.info/lookaround.html)). The syntax for the **end** of the **RegEx** might look something like “($|(?=\s))”.

Let’s see how we would translate the above rules into a **regular expression**:

* First, we need to establish what needs to exist **before** our number. We can’t use \b here, since it includes “-“, which we need in order to match **negative numbers**.   
  Instead, we’ll use a **positive lookbehind**, which will **match** if there’s something **immediately behind** it. We’ll match if we’re either at the **start** of the string (^), or if there’s a **whitespace** **behind** the string:  
  
* Next, we’ll check whether there’s a **hyphen**, signifying a **negative number**:  
  Since having a negative sign **isn’t required**, we’ll use the “?” quantifier, which means “**between 0 and 1 times**”.
* After that, we’ll match any integers – naturally, consisting **one or more digits**:  
  
* Next, we’ll match the **decimal** part of the number, which **might or might not exist** (note: we need to escape the **period** character, as it’s used for something else in RegEx):  
  
* Finally, we’re going to use the same logic for the end of our string as the start – we’re going to match **only** if the number has **either a whitespace or the end of the string (“**$**”)**:  
  

You can follow the table below to help with composing your RegEx:

|  |  |
| --- | --- |
| **Match ALL of these** | **Match NONE of these** |
| 1 -1 123 -123 123.456 -123.456 | 1s s2 s-s -1- \_55\_ s-2 s-3.5 s-1.1 |

Find all the **numbers** from the string and **print them** on the **console**, separated by **spaces**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| 1 -1 1s 123 s-s -123 \_55\_ \_f 123.456 -123.456 s-1.1 s2 -1- zs-2 s-3.5 | 1. -1 123 -123 123.456 -123.456 |

## \*Star Enigma

The war is in its peak, but you, young Padawan, can turn the tides with your programming skills. You are tasked to create a program to **decrypt** the messages of The Order and prevent the death of hundreds of lives.

You will receive several messages, which are **encrypted** using the legendary star enigma. You should **decrypt the messages**, following these rules:

To properly decrypt a message, you should **count all the letters** **[s, t, a, r]** – **case insensitive** and **remove** the count from the **current ASCII value of each symbol** of the encrypted message.

After decryption:

Each message should have a **planet name, population, attack type ('A', as attack or 'D', as destruction) and soldier count.**

The planet name **starts after** **'@'** and contains **only letters from the Latin alphabet**.

The planet population **starts after ':'** and is an **Integer**;

The attack type may be **"A"(attack) or "D"(destruction)** and must be **surrounded by "!"** (exclamation mark).

The **soldier count** starts after **"->"** and should be an Integer.

The order in the message should be: **planet name -> planet population -> attack type -> soldier count.** Each part can be separated from the others by **any character except: '@', '-', '!', ':' and '>'.**

### Input / Constraints

* The **first line** **holds n** – the number of **messages**– **integer in range [1…100];**
* On the next **n** lines, you will be receiving encrypted messages.

### Output

After decrypting all messages, you should print the decrypted information in the following format:

First print the attacked planets, then the destroyed planets.  
"Attacked planets: {attackedPlanetsCount}"  
"-> {planetName}"  
"Destroyed planets: {destroyedPlanetsCount}"  
"-> {planetName}"

The planets should be **ordered by name** **alphabetically.**

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comments** |
| 2  **ST**CDoghudd4=63333$D$0**A**53333  EHf**s**y**ts**nhf?8555&I&2C9555**SR** | Attacked planets: 1  -> Alderaa  Destroyed planets: 1  -> Cantonica | We receive two messages, to decrypt them we calculate the key:  First message has decryption key 3. So we substract from each characters code 3.  **PQ@Alderaa1:30000!A!->20000**  The second message has key 5.  **@Cantonica:3000!D!->4000NM**  **Both messages are valid** and they contain planet, population, attack type and soldiers count.  After decrypting all messages we print each planet according the format given. |
| **Input** | **Output** | **Comments** |
| 3  **tt**(''DG**s**vywge**r**x>6444444444%H%1B9444  GQh**rr**|**A**977777(H(**TTTT**  EHf**s**y**ts**nhf?8555&I&2C9555**SR** | Attacked planets: 0  Destroyed planets: 2  -> Cantonica  -> Coruscant | We receive three messages.  Message one is decrypted with key 4:  **pp$##@Coruscant:2000000000!D!->5000**  Message two is decrypted with key 7:  **@Jakku:200000!A!MMMM**  This is **invalid message**, missing soldier count, so we continue.  The third message has key 5.  **@Cantonica:3000!D!->4000NM** |

"It’s a trap!" – Admiral Ackbar

## 10.\* Rage Quit

Every gamer knows what rage-quitting means. It’s basically when you’re just not good enough and you blame everybody else for losing a game. You press the CAPS LOCK key on the keyboard and flood the chat with gibberish to show your frustration.

Chochko is a gamer, and a bad one at that. He asks for your help; he wants to be the most annoying kid in his team, so when he rage-quits he wants something truly spectacular. He’ll give you **a series of strings followed by non-negative numbers**, e.g. "a3"; you need to print on the console **each string repeated N times**; **convert the letters to uppercase beforehand**. In the example, you need to write back "AAA".

On the output, print first a statistic of the **number of unique symbols** used (the casing of letters is irrelevant, meaning that 'a' and 'A' are the same); the format shoud be "Unique symbols used {0}". Then, **print the rage message** itself.

The **strings and numbers will not be separated by anything**. The input will always start with a string and for each string there will be a corresponding number. The entire input will be given on a **single line**; Chochko is too lazy to make your job easier.

### Input

* The input data should be read from the console.
* It consists of a single line holding a series of **string-number sequences**.
* The input data will always be valid and in the format described. There is no need to check it explicitly.

### Output

* The output should be printed on the console. It should consist of **exactly two lines**.
* On the first line, print the **number of unique symbols used** in the message.
* On the second line, print the **resulting rage message** itself.

### Constraints

* The count of **string-number pairs** will be in the range [1 … 20 000].
* Each string will contain any character **except digits**. The **length** of each string will be in the range [1 … 20].
* The **repeat count** for each string will be an integer in the range [0 … 20].
* Allowed working time for your program: 0.3 seconds. Allowed memory: 64 MB.

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comments** |
| a3 | Unique symbols used: 1  AAA | We have just one string-number pair. The symbol is 'a', convert it to uppercase and repeat 3 times: AAA.  Only one symbol is used ('A'). |
| aSd2&5s@1 | Unique symbols used: 5  ASDASD&&&&&S@ | "aSd" is converted to "ASD" and repeated twice; "&" is repeated 5 times; "s@" is converted to "S@" and repeated once.   1. symbols are used: 'A', 'S', 'D', '&' and '@'. |

'

## 11. \*Letters Change Numbers

Nakov likes Math. But he also likes the English alphabet a lot. He invented a game with numbers and letters from the **English** alphabet. The game was simple. You get a string consisting of a **number between two letters**. Depending on whether the letter was in front of the number or after it you would perform different mathematical operations on the number to achieve the result.

**First** you start with the letter **before** the number.

* If it's **uppercase** you **divide** the number by the letter's **position** in the alphabet.
* If it's **lowercase** you **multiply** the number with the letter's **position** in the alphabet.

**Then** you move to the **letter after** the number.

* If it's **uppercase** you **subtract** its position from the resulted number.
* If it's **lowercase** you **add** its position to the resulted number.

But the game became too easy for Nakov really quick. He decided to complicate it a bit by doing the same but with **multiple** strings keeping track of only the **total sum** of all results. Once he started to solve this with more strings and bigger numbers it became quite hard to do it only in his mind. So he kindly asks you to write a program that **calculates the sum of all numbers after the operations on each number have been done**.

**For example**, you are given the sequence **"**A12b s17G**"**:

We have two strings – "A12b" and "s17G". We do the operations on each and sum them. We start with the letter before the number on the first string. **A is Uppercase** and its position in the alphabet is **1**. So we divide the number 12 with the position 1 **(12/1 = 12)**. Then we move to the letter after the number. **b is lowercase** and its position is 2. So we add 2 to the resulted number **(12+2=14)**. Similarly for the second string **s is lowercase** and its position is 19 so we multiply it with the number **(17\*19 = 323)**. Then we have Uppercase G with position 7, so we subtract it from the resulted number **(323 – 7 = 316)**. Finally, we sum the 2 results and we get **14 + 316=330**.

### Input

The input comes from the console as a **single line, holding the sequence of strings**. Strings are separated by **one or more white spaces**.

The input data will always be valid and in the format described. There is no need to check it explicitly.

### Output

Print at the console a single number: the **total sum of all processed numbers** rounded up to **two digits** after the decimal separator.

### Constraints

The **count** of the strings will be in the range [1 … 10]**.**

* The numbers between the letters will be integers in range [1 … 2 147 483 647]**.**
* Time limit: 0.3 sec. Memory limit: 16 MB.

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comments** |
| A12b s17G | 330.00 | 12/1=12, 12+2=14, 17\*19=323, 323–7=316, **14+316=330** |
| P34562Z q2576f H456z | 46015.13 |  |
| a1A | 0.00 |  |
| P2147483647Z q2576f H456z P34562Z q2576f H456z P34562Z q2576f H456z P34562Z |  |  |

## \*SoftUni Bar Income

Let`s take a break and visit the game bar at SoftUni. It is about time for the people behind the bar to go home and you are the person who has to draw the line and calculate the money from the products that were sold throughout the day. Until you receive a line with text "end of shift" you will be given lines of input. But before processing that line you have to do some validations first.

**Each valid order** should have a **customer, product, count and a price:**

* Valid customer's name should be **surrounded by '%'** and must **start with a capital letter**, followed by **lower-case letters**
* Valid product **contains any word character** and must be **surrounded by '<' and '>'**
* Valid count is an **integer**, **surrounded by '|'**
* Valid price is any **real number followed by '$'**

The parts of a valid order should appear in the order given: **customer, product, count and a price**.

Between each part there can be other symbols, except (**'|', '$', '%' and '.'**)

For each valid line print on the console: "{customerName}: {product} - {totalPrice}"

When you receive "end of shift" print the total amount of money for the day **rounded to 2 decimal places** in the following format:"Total income: {income}"**.**

### Input / Constraints

* Strings that you have to process until you receive text "end of shift".

### Output

* Print all of the valid lines in the format "{customerName}: {product} - {totalPrice}"
* After receiving "end of shift" print the total amount of money for the day rounded to 2 decimal places in the following format: "Total income: {income}"
* Allowed working **time** / **memory**: **100ms** / **16MB**.

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comment** |
| %George%<Croissant>|2|10.3$  %Peter%<Gum>|1|1.3$  %Maria%<Cola>|1|2.4$  end of shift | George: Croissant - 20.60  Peter: Gum - 1.30  Maria: Cola - 2.40  Total income: 24.30 | Each line is valid, so we print each order, calculating the total price of the product bought.  At the end we print the total income for the day |
| %InvalidName%<Croissant>|2|10.3$  %Peter%<Gum>1.3$  %Maria%<Cola>|1|2.4  %Valid%<Valid>valid|10|valid20$  end of shift | Valid: Valid - 200.00  Total income: 200.00 | On the first line, the customer name isn`t valid, so we skip that line. The second line is missing product count.  The third line don`t have a valid price.  And only the forth line is valid |