# Exercise: Text Processing and Regular Expressions

Problems for exercises and homework for the ["Technology Fundamentals" course @ SoftUni](https://softuni.bg/courses/technology-fundamentals).

You can check your solutions in [Judge.](https://judge.softuni.bg/Contests/1331)

## 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** |  | **Input** | **Output** |  | **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**.

### Input

You will receive **single** **line** with the string.

### Output

Print what is left from the string after 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** of **1**. 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 now with strength of 2. We delete the digit and we find **another** explosion. At this point the string looks like this: abv>>>>2asdasd.  4th explosion is with strength **2**. 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 |  |

# Regular Expressions

## Extract Emails

Write a program to **extract all email addresses from a given text**. The text comes at the only input line. Print the emails on the console, each at a separate line. Emails are considered to be in format <user>@<host>, where:

* **<user>** is a sequence of **letters** and **digits**, where '.', '-' and '\_' can appear between them.
  + Examples of valid users: "**stephan**", "**mike03**", "**s.johnson**", "**st\_steward**", "**softuni-bulgaria**", "**12345**".
  + Examples of invalid users: ''**--123**", "**.....**", "**nakov\_-**", "**\_steve**", "**.info**".
* **<host>** is a sequence of at least two words, separated by dots '**.**'. Each word is sequence of letters and can have hyphens '**-**' between the letters.
  + Examples of hosts: "**softuni.bg**", "**software-university.com**", "**intoprogramming.info**", "**mail.softuni.org**".
  + Examples of invalid hosts: "**helloworld**", "**.unknown.soft.**", "**invalid-host-**", "**invalid-**".
* Examples of **valid emails**: **info@softuni-bulgaria.org**, **kiki@hotmail.co.uk**, **no-reply@github.com**, **s.peterson@mail.uu.net**, **info-bg@software-university.software.academy**.
* Examples of **invalid emails**: **--123@gmail.com**, **…@mail.bg**, **.info@info.info**, **\_steve@yahoo.cn**, **mike@helloworld**, **mike@.unknown.soft.**, **s.johnson@invalid-**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| Please contact us at: support@github.com. | support@github.com |
| Just send email to s.miller@mit.edu and j.hopking@york.ac.uk for more information. | s.miller@mit.edu  j.hopking@york.ac.uk |
| Many users @ SoftUni confuse email addresses. We @ Softuni.BG provide high-quality training @ home or @ class. –- steve.parker@softuni.de. | steve.parker@softuni.de |

## Furniture

Write a program to calculate the total cost of different types of furniture. You will be given some lines of input until you receive the line "Purchase". For the line to be valid it should be in the following format:

**">>{furniture name}<<{price}!{quantity}"**

The price can be floating point number or whole number. Store the names of the furniture and the total price. At the end print the each bought furniture on separate line in the format:

**"Bought furniture:**

**{1st name}**

**{2nd name}**

**…"**

And on the last line print the following: **"Total money spend: {spend money}"** formatted to the second decimal point.

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comment** |
| >>Sofa<<312.23!3  >>TV<<300!5  >Invalid<<!5  Purchase | Bought furniture:  Sofa  TV  Total money spend: 2436.69 | Only the Sofa and the TV are valid, for each of them we multiply the price by the quantity and print the result |

## Race

Write a program that processes information about a race. On the **first line** you will be given **list of participants** **separated by ", "**. On the next few lines until you receive a line **"end of race"** you will be given some info which will be some **alphanumeric characters**. In between them you could have some **extra characters which you should ignore**. For example: **"G!32e%o7r#32g$235@!2e"**. The **letters are the name** of the person and the **sum of the digits is the distance** he ran. So here we have **George** who ran **29 km**. Store the information about the person only **if the list of racers contains the name of the person**. If you receive the **same person more than once just add the distance to his old distance**. At the end **print the top 3 racers** ordered by **distance in descending** in the format:

**"1st place: {first racer}**

**2nd place: {second racer}**

**3rd place: {third racer}"**

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comment** |
| George, Peter, Bill, Tom  G4e@55or%6g6!68e!!@  R1@!3a$y4456@  B5@i@#123ll  G@e54o$r6ge#  7P%et^#e5346r  T$o553m&6  end of race | 1st place: George  2nd place: Peter  3rd place: Tom | On the 3rd input line we have Ray. He is not in the list, so we do not count his result. The other ones are valid. George has total of 55 km, Peter has 25 and Tom has 19. We do not print Bill because he is on 4th place. |

## \*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 |

## \*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

## \*Nether Realms

Mighty battle is coming. In the stormy nether realms, demons are fighting against each other for supremacy in a duel from which only one will survive.

Your job, however is not so exciting. You are assigned to **sign in all the participants** in the nether realm's mighty battle's demon book, which of course is **sorted alphabetically**.

A demon's **name contains his health and his damage**.

The **sum of the asci codes** of **all characters** (excluding numbers (0-9), arithmetic symbols (**'+', '-', '\*', '/'**) and delimiter dot (**'.'**)) gives a **demon's total health**.

**The sum of all numbers** in his name forms his base damage. Note that you should consider the plus **'+'** and minus **'-'** signs (e.g. **+10 is 10** and **-10 is -10**). However, there are some symbols (**'\*'** and **'/'**) that can further **alter the base damage by multiplying or dividing it by 2** (e.g. in the name "m**15**\*/c**-5.0**", the base damage is **15 + (-5.0) = 10** and then you need to multiply it by 2 (e.g. 10 \* 2 = 20) and then divide it by 2 (e.g. 20 / 2 = 10)).

So, **multiplication and division** are applied **only after all numbers are included** in the calculation and **in the order they appear in the name**.

You will get all demons **on a single line**, separated by commas and zero or more blank spaces. Sort them in **alphabetical order** and print their names **along their health and damage**.

### Input

The input will be read from the console. The input consists of a **single line** containing all demon names **separated by commas and zero or more spaces** in the format: **"{demon name}, {demon name}, … {demon name}"**

### Output

Print all demons **sorted by their name in ascending order**, each on a separate line in the format:

* **"{demon name} - {health points} health, {damage points} damage"**

### Constraints

* A demon's name will contain **at least one character**
* A demon's name **cannot contain** blank spaces ' ' or commas ','
* A **floating point number will always have digits before and after its decimal separator**
* **Number** in a demon's name **is considere**d everything that is a valid integer or floating point number (with dot '.' used as separator). For example, all these are valid numbers: '4', '+4', '-4', '3.5', '+3.5', '-3.5'

### Examples

|  |  |  |
| --- | --- | --- |
| **Input** | **Output** | **Comments** |
| M3ph-0.5s-0.5t0.0\*\* | M3ph-0.5s-0.5t0.0\*\* - 524 health, 8.00 damage | M3ph-0.5s-0.5t0.0\*\*:  Health = 'M' + 'p' + 'h' + 's' + 't' = 524 health.  Damage = (3 + (-0.5) + (-0.5) + 0.0) \* 2 \* 2 = 8 damage. |
| M3ph1st0\*\*, Azazel | Azazel - 615 health, 0.00 damage  M3ph1st0\*\* - 524 health, 16.00 damage | Azazel:  Health - 'A' + 'z' + 'a' + 'z' + 'e' + 'l' = 615 health. Damage - no digits = 0 damage.  M3ph1st0\*\*:  Health - 'M' + 'p' + 'h' + 's' + 't' = 524 health.  Damage - (3 + 1 + 0) \* 2 \* 2 = 16 damage. |
| Gos/ho | Gos/ho - 512 health, 0.00 damage |  |

## \*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** | **Comment** |
| 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 |  |