# Exercise: Regular Expressions

Problems for exercise and homework for the [Python Fundamentals Course @SoftUni](https://softuni.bg/trainings/2442/python-fundamentals-september-2019). Submit your solutions in the SoftUni judge system at <https://judge.softuni.bg/Contests/1743>

## Capture the Numbers

Write a program that **finds all numbers** in a sequence of strings.

The **output** is all the numbers, **extracted** and **printed on a single line** – each separated by a **single space**.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| The300  What is that?  I think it's the 3rd movie  Lets watch it at 22:45 | 300 3 22 45 |
|  |  |
| **Input** | **Output** |
| 123a456  789b987  654c321  0 | 123 456 789 987 654 321 0 |
| **Input** | **Output** |
| Let's go11!!!11!  Okey!1! | 11 11 1 |

## Find Variable Names in Sentences

Write a program that finds all **variable names** in a given string. A variable name starts with an **underscore** ("\_") and contains **only Capital and Non-Capital English Alphabet letters and digits**. Extract only their names, **without the underscore.** Try to do this **only with regular expressions**.

The **output** consists of all variable names, **extracted** and **printed on a single line**, each **separated** by a **comma**.

|  |  |
| --- | --- |
| **Input** | **Output** |
| The \_id and \_age variables are both integers. | id,age |
|  |  |
| **Input** | **Output** |
| Calculate the \_area of the \_perfectRectangle object. | area,perfectRectangle |
| **Input** | **Output** |
| \_\_invalidVariable \_evenMoreInvalidVariable\_ \_validVariable | validVariable |

## Find Occurrences of Word in Sentence

Write a program that finds, **how many times** a **given word**, is **used** in a **given sentence.** Note that letter case does not matter – it is **case-insensitive**.

The **output** is a single number indicating the **amount of times** the sentence contains the word.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| The waterfall was so high, that the child couldn't see its peak.  the | 2 |
|  |  |
| **Input** | **Output** |
| How do you plan on achieving that? How? How can you even think of that?  how | 3 |
| **Input** | **Output** |
| There was one. Therefore I bought it. I wouldn't buy it otherwise.  there | 1 |

## 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 |

## \*Extract the Links

Write a program that **extracts links** from a **given text**. The text will come in the form of strings, each representing a sentence. You need to extract **only the valid links** from it. Example:

"www.internet.com"

**Sub-Domain**  **Domain** **name** **Domain extension**

The **Sub-Domain** must always be "www". The **Domain name** can consist of English alphabet letters (**uppercase** and **lowercase**), digits and dashes ("–"). The **Domain extension** consists of one or more **domain blocks**, a **domain block** consists of a **dot** followed by **one or more lowercase** English alphabet **letters**, a **Domain extension** must have at least **one** **domain block** in order to be **valid**. The Sub-Domain and Domain name must be separated by a single **dot**. Any link that **does NOT follow** the specified above rules should be treated as **invalid**.

**Example incorrect links:**

* "**ww**.justASite.bg"
* "**lel**.awesome.com"
* "www.stamat**\_**gosho.hit.bg"
* "www.no-symb**#^**ols-allow**%**ed.com"
* "www.pesho**.12**"
* "www.gosho-site**.**"
* "www.example-site**.\_\*^#**"

**Example correct links:**

* "Some textwww.softuni.bg"
* "Just a link in a www.sea-of-text.bg-swimming around"
* "Instruments www.Instruments.rom.com.trombone2000 Instrument here"
* "All your ice cream flavors-www.scream.for.ice.cream(We also have squirrels)"

You will receive strings until the command **"End"**

The **output** is all valid links you've found, printed – each on a new line.

### Examples

|  |  |
| --- | --- |
| **Input** | **Output** |
| Join WebStars now for free, at [www.web-stars.com](http://www.web-stars.com)  You can also support our partners:  Internet - [www.internet.com](http://www.internet.com)  WebSpiders - [www.webspiders101.com](http://www.webspiders101.com)  Sentinel - [www.sentinel.-ko](http://www.sentinel.-ko)  End | www.web-stars.com  www.internet.com  www.webspiders101.com |
|  |  |
|  |  |
| **Input** | **Output** |
| Need information about cheap hotels in London?  You can check us at [www.london-hotels.co.uk](http://www.london-hotels.co.uk)!  We provide the best services in London.  Here are some reviews in some blogs:  London Hotels are awesome! - [www.indigo.bloggers.com](http://www.indigo.bloggers.com)  I am very satisfied with their services - ww.ivan.bg  Best Hotel Services! - [www.rebel21.sedecrem.moc](http://www.rebel21.sedecrem.moc)  End | www.london-hotels.co.uk  www.indigo.bloggers.com  www.rebel21.sedecrem.moc |

## \*Secret Data

Write a program that hides essential client data from secret documents that went public. You have to hide people's names, phone numbers, ids and secret base names.

* The **names of the clients** will be preceded by a single **asterisk** ("\*"), they will always be **exactly one word**, they will contain **only English alphabet letters**, they will **start with a capital letter** and they will always be followed by either a **space,** a **tabulation** or the **end of the string**. Anything else is **NOT** **to be considered** as a name.
* The **phone numbers** of the clients will be preceded by a single **plus sign** ("+") and will consist of exactly 10 symbols. The phone numbers can consist only of **digits** and **dashes** and they will always be followed by a **space, tabulation** or the **end of the string**. Anything else is **NOT to be considered** as a phone number.
* The **ID**s of the clients will be preceded by a single **exclamation mark** ("!"). The IDs of the clients will consist only of **Lowercase** and **Uppercase English alphabet letters** and **digits** and they will always be followed by a **space, tabulation** or the **end of the string**. Anything else is **NOT to be considered** as an ID.
* The **names of the secret bases** will be preceded by a single **underscore** ("\_") and will consist only of **Uppercase** and **Lowercase English alphabet letters** and **digits** and they will always be followed by a **space, tabulation** or the **end of the string**. Anything else is **NOT to be considered** as a secret base name.

### Constraints

* **Usernames, phone numbers, IDs and names of secret bases** can start glued to other text.
* **Usernames, phone numbers, IDs and names of secret bases** will never be split given across 2 lines.
* **Usernames, phone numbers, IDs and names of secret bases** will always have a **space, tabulation** or the **end of the string** after them.

The **input** comes as strings on separate lines until the command **"End"**. Each string represents a sentence of the secret document. You need to find every **client data element** that is supposed to be secret, and **hide it**, by **replacing it** with a **string of** **vertical bars** ("|") with the **same length**, as the **discovered element**.

**NOTE: The preceding symbol counts towards the discovered element's length when deciding how many pipes to use to cover it.**

The **output** should be the same document, with the sensitive **client** **data replaced by pipes**. See the example for more info.

### Example

|  |  |
| --- | --- |
| **Input** | **Output** |
| Agent \*Ivankov was in the room when it all happened  The person in the room was heavily armed.  Agent \*Ivankov had to act quick in order.  He picked up his phone and called some unknown number  End | Agent |||||||| was in the room when it all happened.  The person in the room was heavily armed.  Agent |||||||| had to act quick in order.  He picked up his phone and called some unknown number. |