Link Ref: <https://regexr.com/>

<https://viblo.asia/p/hoc-regular-expression-va-cuoc-doi-ban-se-bot-kho-updated-v22-Az45bnoO5xY>

Cú pháp để bắt chuỗi, vào thao tác với chuỗi

**Trước hết các bạn hãy đọc qua một lượt các quy tắc bắt chuỗi (matching) đơn giản nhất của RegEx**

* [xyz] Tìm và so sánh tất cả ký tự nằm trong dấu ngoặc vuông và trùng khớp với 1 ký tự trong dấu ngoặc vuông. Ví dụ: [31] sẽ trùng khớp với **3** hoặc **1**, [0123456789] sẽ trùng khớp với bất kỳ một ký tự nào trong khoảng từ **0** đến **9**.
* [a-z] So sánh và trùng khớp với một ký tự nằm trong khoảng chỉ định. Ví dụ: [a-z] sẽ trùng khớp với một ký tự trong khoảng từ a đến z nằm trong chuỗi cần test. [0-9] sẽ trùng khớp với bất kỳ một ký tự nào trong khoảng từ **0** đến **9**.
* [^xyz] So sánh và không trùng khớp với những ký tự nằm trong khoảng chỉ định. Dấu ^ (dấu mũ) nằm trong dấu ngoặc vuông là một dấu phủ định. Ví dụ: [^a-z] sẽ không trùng khớp với tất cả các ký tự nằm trong khoảng từ **a** đến **z**.
* ^ Trùng khớp với phần đầu của chuỗi đích. Ví dụ: ^a sẽ trùng khớp với chữ **a** trong chuỗi **abc**, ^\w+ sẽ trùng khớp với chữ đầu tiên – chữ **"the"** của chuỗi **"The quick brown fox jumps over the lazy dog"**.
* $ Trùng khớp với phần cuối của chuỗi đích. Ví dụ: c$ sẽ trùng khớp với chữ c trong chuỗi abc, \w+$ sẽ trùng khớp với chữ cuối – chữ **"dog"** của chuỗi **"The quick brown fox jumps over the lazy dog"**.
* + Trùng khớp với 1 hoặc nhiều lần ký tự đứng trước nó. Ví dụ \d+ sẽ chỉ trùng với chuỗi có từ 1 con số trở lên.
* \* Trùng khớp với 0 hoặc nhiều lần ký tự đứng trước nó. Ví dụ \d\* sẽ trùng với chuỗi có chứa 1 chữ số hoặc k có chữ số nào cũng đc.
* ? Trùng khớp với 0 hoặc 1 lần ký tự đứng trước nó. Tương tự như \* nhưng nó lại chỉ nhân lên 1 lần. \* thì nhân lên nhiều lần.
* . Trùng khớp với 1 ký tự đơn bất kỳ ngoại trừ ký tự ngắt dòng (line-break) và cũng không lấy được ký tự có dấu (unicode). Ví dụ: . sẽ trùng khớp với ký tự **a hoặc b hoặc c** trong chuỗi **abc**. Nhưng . sẽ không bắt được các chữ ă hoặc ê.
* x{n} Trùng khớp đúng với n lần ký tự đứng trước nó. n là một số không âm. Ví dụ \d{2} sẽ bắt đc các số có 2 chữ số đứng liền nhau.
* x{n,} Trùng khớp với ít nhất n lần ký tự đứng trước nó. n là một số không âm.Ví dụ \d{2,} sẽ bắt đc các số có từ 2 chữ số trở lên đứng liền nhau.
* x{n,m} Trùng khớp với ít nhất n lần và nhiều nhất là m lần ký tự đứng trước nó. n và m là một số không âm và n <= m. Ví dụ: a{1,3} sẽ khớp với **hah, haah, haaah** nhưng không khớp với **haaaah**.
* x|y Trùng khớp với x hoặc y. Ví dụ: slow|fast sẽ khớp với chữ **slow** hoặc **fast** trong chuỗi đích.
* \b Trùng khớp với toàn bộ ký tự đứng trước nó. Ví dụ: hello\b sẽ trùng khớp với toàn bộ từ hello trong chuỗi **hello world** nhưng sẽ không khớp với chuỗi **helloworld**.
* \B Ngược lại với \b, \B sẽ không khớp với toàn bộ mà chỉ 1 phần ký tự đứng trước nó. Ví dụ: hello\B sẽ trùng khớp với chữ hello trong chuỗi **helloworld** nhưng sẽ không khớp với chuỗi **hello world**.
* \d Trùng khớp 1 ký tự số (digit).
* \D Trùng khớp 1ký tự không phải số (non-digit).
* \s Trùng khớp 1 ký tự khoảng trắng (whitespace) bao gồm khoảng trắng tạo ra bởi phím Tab.
* \S Trùng khớp với 1 ký tự không phải là khoảng trắng (non-whitespace).
* \w Trùng khớp với các ký tự là từ (word) bao gồm dấu \_ (underscore) và chữ số.
* \W Trùng khớp với các ký tự không phải là từ (non-word). Ví dụ: \W sẽ khớp với ký tự % trong chuỗi **"100%"**.
* \uxxxx Trùng khớp với 1 ký tự unicode. Ví dụ: \u00FA sẽ khớp với ký tự **"ú"**, \u00F9 sẽ khớp với ký tự **"ù"**.
* \pL Trùng khớp với một ký tự Unicode bất kỳ ngoại trừ dấu cách. Đây chính là cú pháp viết hoàn hảo hơn của dấu .,Ví dụ \pL+ sẽ lấy được chuỗi truyền, thuyết trong chuỗi **"truyền thuyết"**.