Regex Engine Performance

Regex Performance

Poorly written pattern may work well for perfect-match

Worst-case performance is seen for partial or no-match scenarios

Good: $O(n)$, $O(n^2)$

Bad: $O(2^n)$, $O(3^n)$

where, n is the number of characters in the input text

Exponential Runtime Example

```
Problem: Match a word
Pattern: ^(\w*)*$
Text (match):           12345678901234567890
Text (partial match):   12345678901234567890!
```

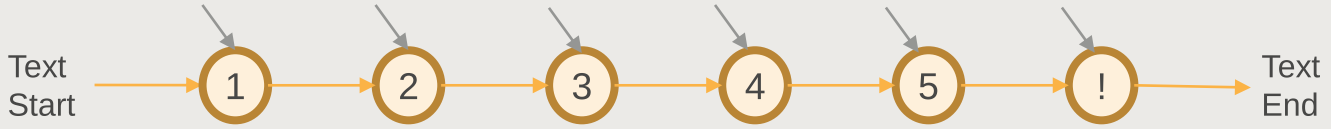
\w* => capture zero or more-word characters

(\w*)* => capture 0 or more groups

Pattern works fine for positive match. For a partial match, the performance degrades rapidly, and every additional character doubles the response time.

Issue

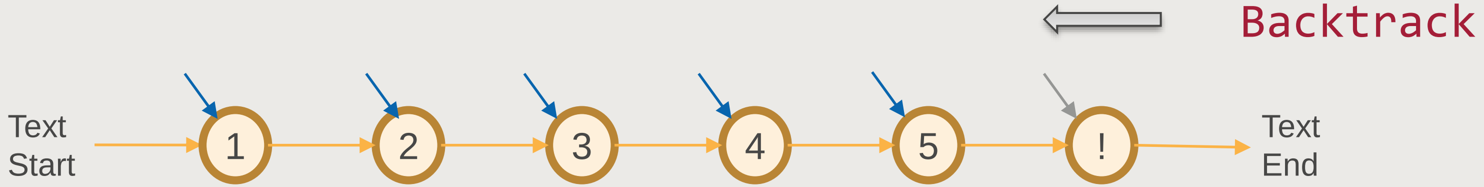
```
Pattern: ^(\w*)*$
```



```
\w*: 1
\w*: 12
\w*: 123
\w*: 1234
\w*: 12345
\w*: 12345! Does not match \w and end of string $ match fails.
```

Exponential Degradation

Pattern: ^(\w*)*$



```
Backtrack
(\w*)*: (1234)(5)!
(\w*)*: (123)(45)! => (123)(4)(5)!
(\w*)*: (12)(345)! => (12)(34)(5) => (12)(3)(45) => (12)(3)(4)(5)
(\w*)*: (1)(2345) => (1)(2)(345) => (1)(2)(3)(45)=>(1)(2)(3)(4)(5) => (1)(23)(45) =>
```

Solution

Pattern: `(\w*)*   => (\w*)(\w*)(\w*)…`

Multiple similar greedy patterns may cause performance issues

Option 1: Remove group level quantifier

```
  ^(\w*)*$  => ^(\w*)$
               ^\w*$  or ^(?:\w*)$  (disable group capture)
```

Option 2: Precise terminating condition. Every word in group should end in a word boundary.

```
  ^(\w*)*$  => ^(\w*\b)*$
```

Regex Compiled Objects

Two-ways to use Regex

Directly invoke `re` module methods

Compile pattern and invoke methods of compiled object

- Additional parameters such as start position, end position
- Consistent performance
- Control over object lifetime

re Module

Internally compiles and caches patterns

In most cases performance is similar to compiled object

Limited cache-size – clears entire cache when full

- Python 2 => 100 objects
- Python 3 => 512 objects
- Increased latency when cache is flushed and rebuilt

Chandra Lingam

60,000+ Students

aws certified Solutions Architect Associate

aws certified Developer Associate

aws certified SysOps Administrator Associate

aws certified Solutions Architect Professional

aws certified Machine Learning Specialty

For AWS self-paced video courses, visit:

https://www.cloudwavetraining.com/

Cloud Wave LLC