

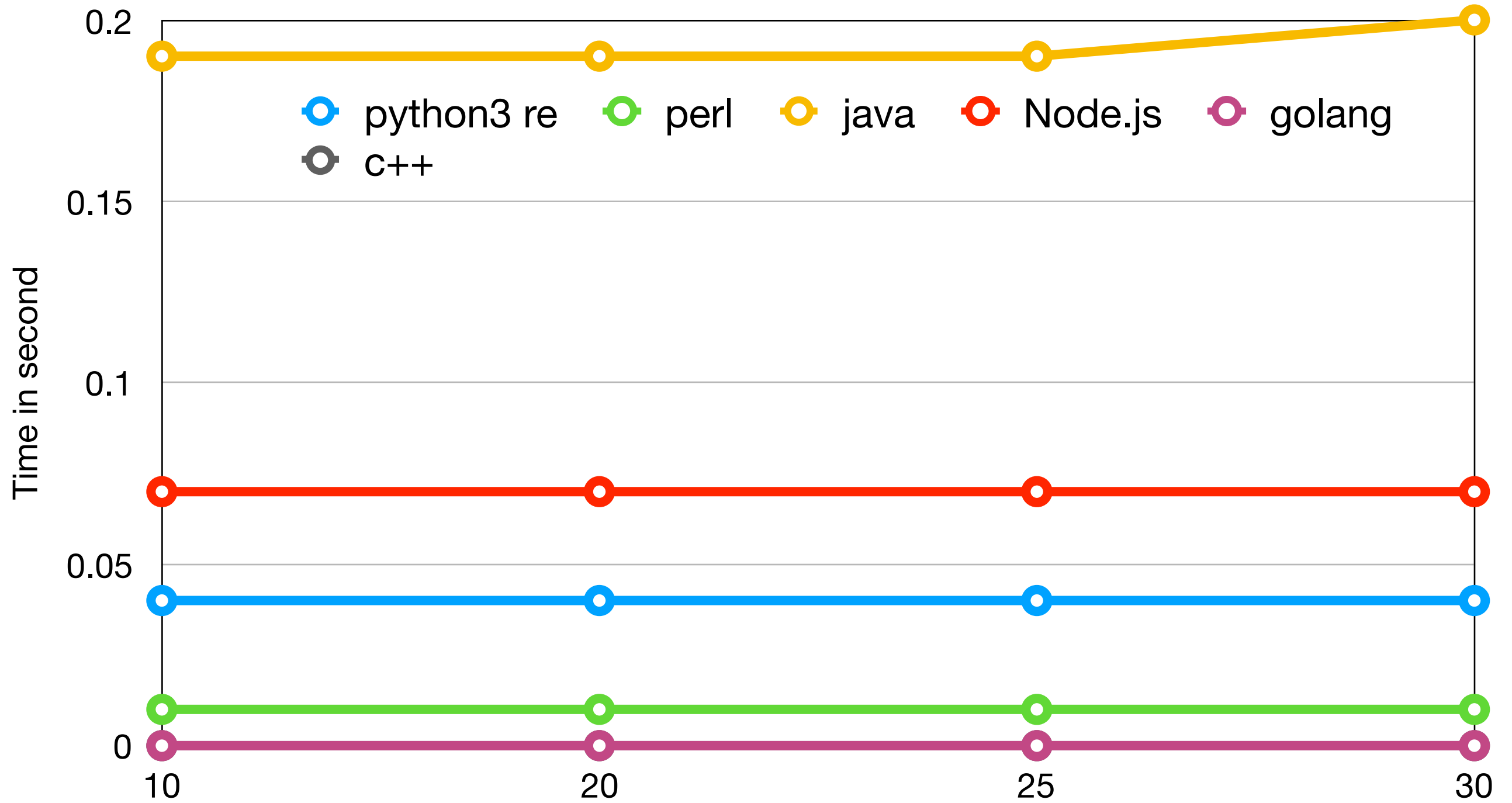
Regex

403410033 資工四 曾俊宏

Fun fact

- Test 1
 - Regex: `a?a?a?a?a?a?a?a?a?aaaaaaaaaaa`
 - Input: `aaaaaaaaaaaaaaaaaaaaa`
 - Key point: all `?` has a matching `a`
- Test 2
 - Regex: `a?a?a?a?a?a?a?a?a?aaaaaaaaaaa`
 - Input: `aaaaaaaaaaa`
 - Key point: all `?` has no matching `a`

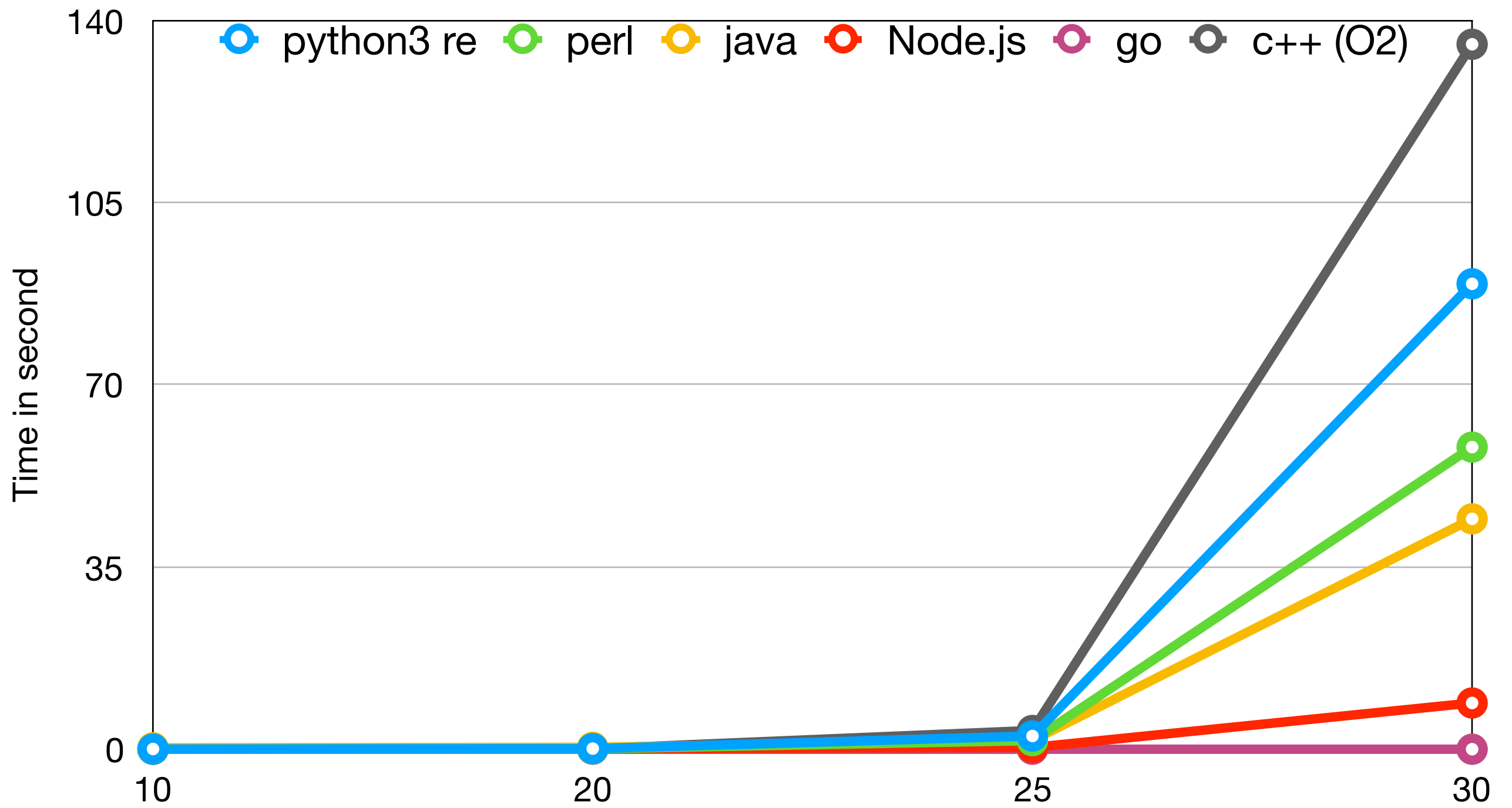
Test 1



So far so good

- Regex `a?a?a?a?aaaaa` matching `aaaaaaaaaaa`
- When length = 30, all programming languages finished matching in under 0.2 second

Test 2



WTF?

- Regex `a?a?a?a?aaaaa` matching `aaaaa`
- When length = 30, all programming languages except Golang has exponential matching time!
 - Why Golang survived?
 - Russ Cox in Golang project implemented regex without recursive backtracking
 - Pointed out the issue in 2007, still, no improvements for all programming languages he mentioned so far

Specification

- 目前：
 - 支援 $?^*|+$
 - 用 NFA
- 目標：
 - 支援 escape sequences, e.g. `\(, \\, etc.`
 - 支援 character classes, e.g. `[0-9], \w, etc.`
 - 用 DFA

Procedure

- Regex to postfix
 - 搞死了，資料結構沒學好
- Postfix to NFA
 - 很容易有dangling pointer炸開的bug
- Matching!
 - 目前是跟著報告實作，所以 code 速度是跟 golang 差不多
 - 但是目前有些帶有 | 的 pattern 會炸開 -> null pointer...