Steaming Data

One Record at a Time

miki (atebeka

CEO



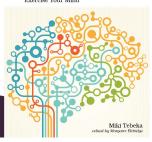


Effective Go Recipes Fast Solutions to Common Tasks

The Pragmatic Programmers



Go Brain Teasers





Python Brain Teasers





Pandas Brain Teasers







Range Over Function

Go 1.23, August 2024

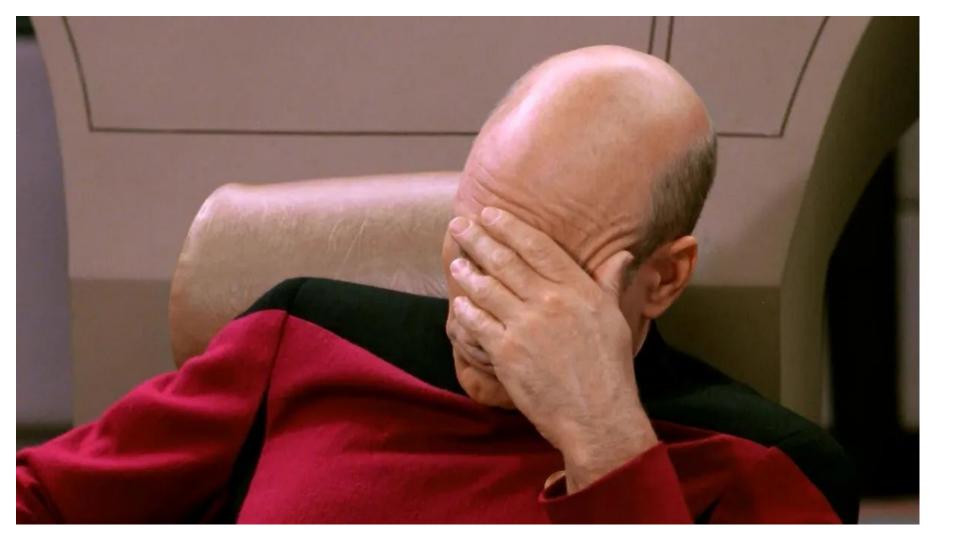
Generators

Python 2.3, July 2003

cmd/tick/tick.go

Load logs from disk so we can query them.

```
$ head logs/1/x00
199.72.81.55 - - [01/Jul/1995:00:00:01
-0400] "GET /history/apollo/ HTTP/1.0" 200
6245
unicomp6.unicomp.net - -
[01/Jul/1995:00:00:06 -0400] "GET
/shuttle/countdown/ HTTP/1.0" 200 3985
199.120.110.21 - - [01/Jul/1995:00:00:09
-0400] "GET
/shuttle/missions/sts-73/mission-sts-73.htm
1 HTTP/1.0" 200 4085
```







Write a Go regexp to parse the following log lines:

199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET

/history/apollo/ HTTP/1.0" 200 6245

unicomp6.unicomp.net - - [01/Jul/1995:00:00:06

-0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985

199.120.110.21 - - [01/Jul/1995:00:00:09 -0400] "GET

/shuttle/missions/sts-73/mission-sts-73.html

HTTP/1.0" 200 4085

burger.letters.com - - [01/Jul/1995:00:00:11 -0400]

"GET /shuttle/countdown/liftoff.html HTTP/1.0" 304 0

199.120.110.21 - - [01/Jul/1995:00:00:11 -0400] "GET

/shuttle/missions/sts-73/sts-73-patch-small.gif

HTTP/1.0" 200 4179

burger.letters.com - - [01/Jul/1995:00:00:12 -0400]

"GET /images/NASA-logosmall.gif HTTP/1.0" 304 0

burger.letters.com - - [01/Jul/1995:00:00:12 -0400]

Reply to Claude...

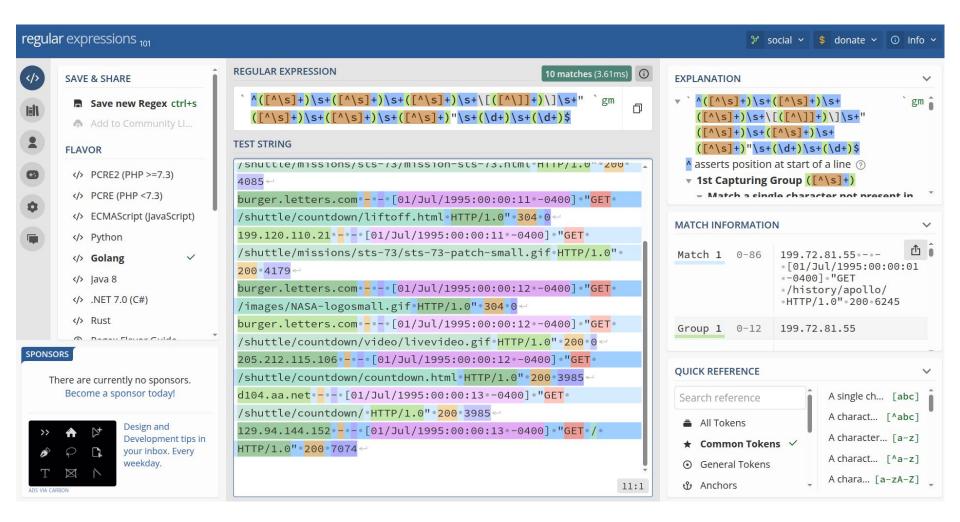


Claude 3.7 Sonnet ~



```
// LogEntry represents a parsed log line
type LogEntry struct {
  RemoteHost
               string
  Identity
               string
 UserTD
               string
  Timestamp
                string
  Method
                string
  Path
                string
  Protocol
                string
  StatusCode
               string
  ResponseBytes string
func main() {
  // Example log line for testing
  logLine := `199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "
  // Define the regular expression pattern
  pattern := ^([^{s}+)\s+([^{s}+)\s+([^{s}+)\s+([^{s}+)\])
  re := regexp.MustCompile(pattern)
```

// Parse log line using regexp





parser/parser.go

loader/simple

cmd/load

\$ du -sh logs 1.2G logs



loader/seq

Map

lower: ["Suite", "Up"] -> ["suite", "up"]

Filter

 \geq 0: [2, -1, 3, 4] -> [2, 3, 4]

Reduce

sum: $[1, 2, 3, 4] \rightarrow 10$

```
reduce map
SELECT SUM(vol * price)
FROM trades
WHERE symbol = 'IBM';
             filter
```

loader/mapper

Load logs from disk so we can query them.





Fast

Accelerated expression evaluation in performance-critical paths from nanoseconds to microseconds.



Portable

Developer friendly, light weight with common syntax across multiple Google and external systems.



Extensible

Supports subsetting and extension, easy to embed and tailor to configuration and policy requirements.



Safe

Non-Turing complete, and only accesses data provided by the host application.



github.com/google/cel-go

Transfer-Encoding: chunked

jsonlines ndjson

cmd/httpd



Questions?



github.com/tebeka/talks/tree/master/streams