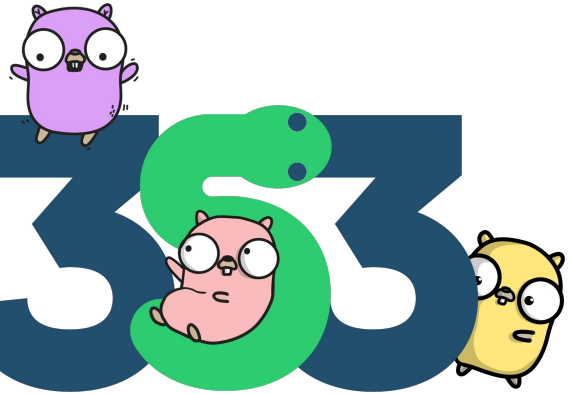
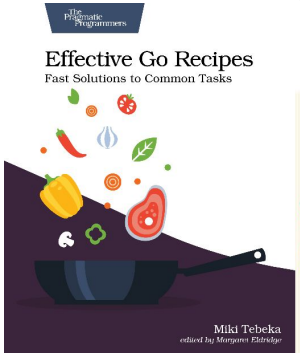
Steaming Data

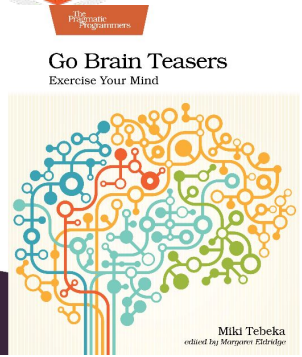
One Record at a Time

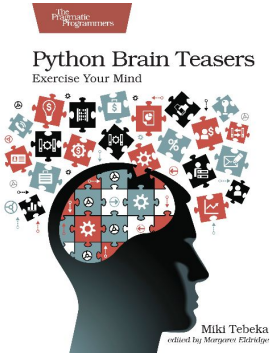
miki @tebeka

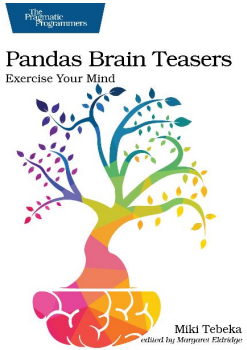
CEO,
CTO,
UFO











GopherCON Israel

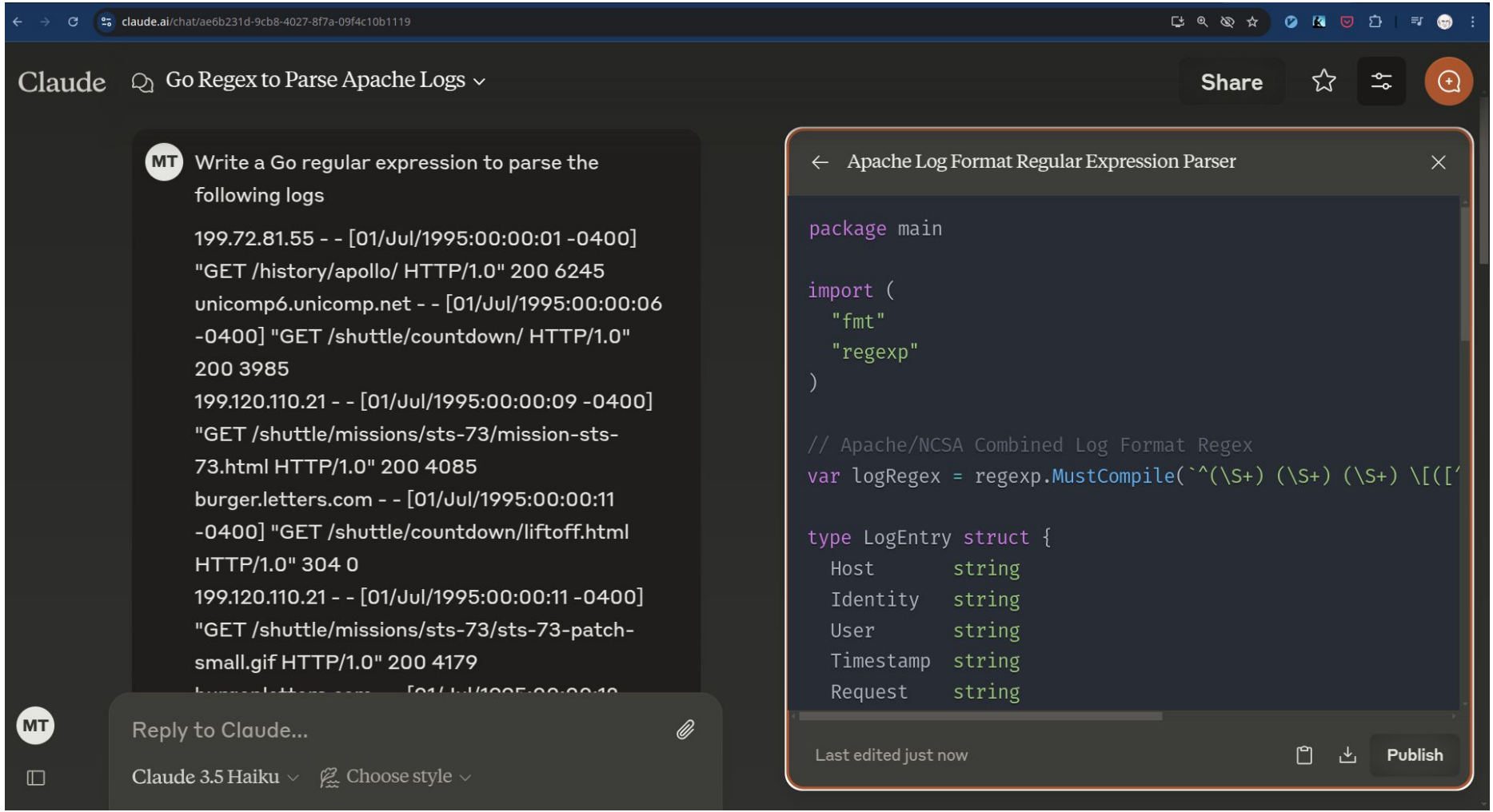
info@gophercon.org.il

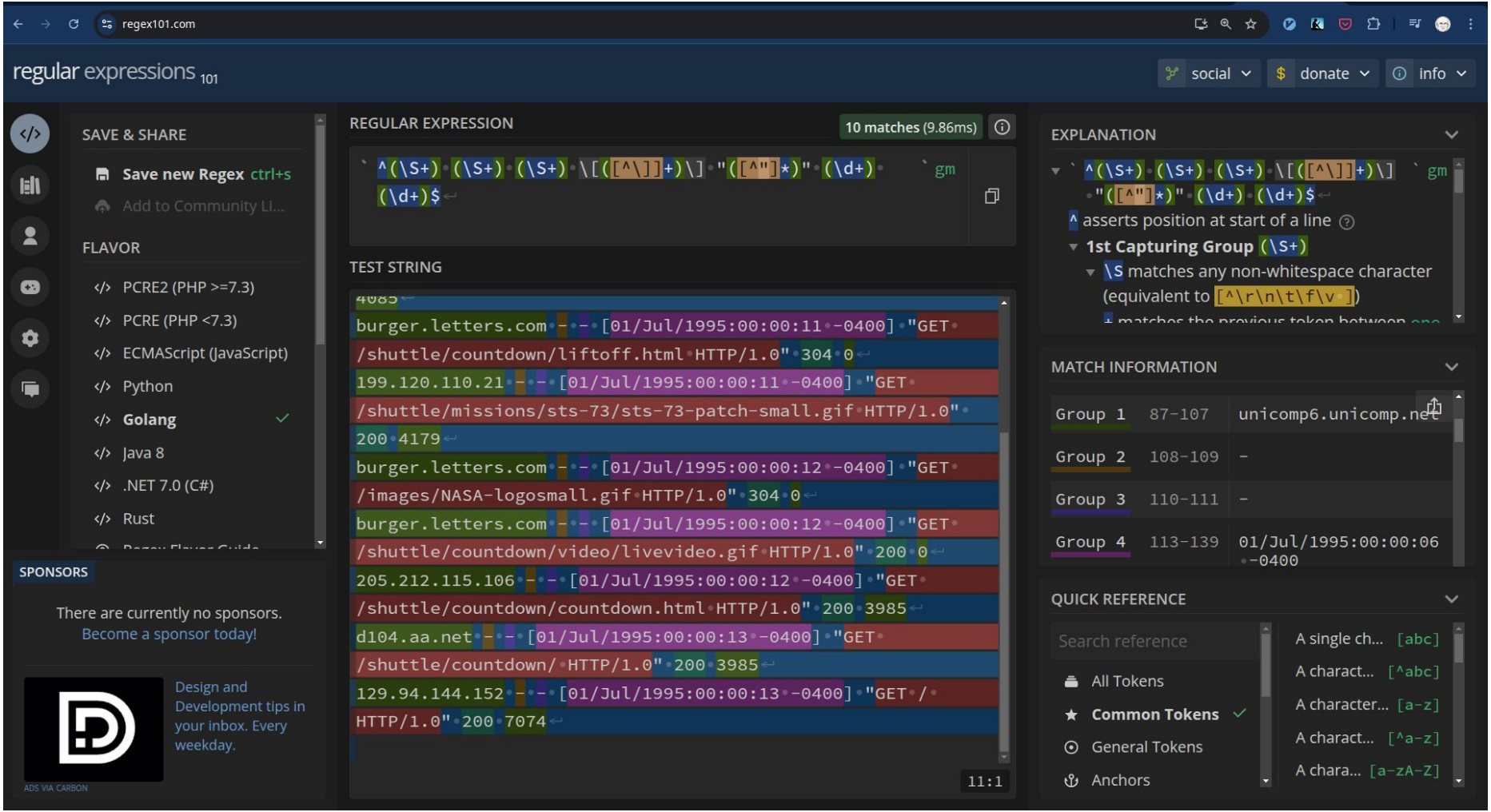
CFP

SPONSOR

Load logs from file 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
l HTTP/1.0" 200 4085
```


claude.ai/chat/ae6b231d-9cb8-4027-8f7a-09f4c10b1119
Claude
Go Regex to Parse Apache Logs
Share
MT
Write a Go regular expression to parse the following logs
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
Reply to Claude...
Claude 3.5 Haiku
Choose style
Apache Log Format Regular Expression Parser
package main
import (
"fmt"
"regexp"
)
// Apache/NCSA Combined Log Format Regex
var logRegex = regexp.MustCompile(`^(\S+) (\S+) (\S+) \[([
type LogEntry struct {
Host string
Identity string
User string
Timestamp string
Request string
Last edited just now
Publish


regex101.com
regular expressions 101
social
donate
info
SAVE & SHARE
Save new Regex ctrl+s
Add to Community Li...
FLAVOR
PCRE2 (PHP >=7.3)
PCRE (PHP <7.3)
ECMAScript (JavaScript)
Python
Golang
Java 8
.NET 7.0 (C#)
Rust
SPONSORS
There are currently no sponsors.
Become a sponsor today!
Design and Development tips in your inbox. Every weekday.
ADS VIA CARBON
REGULAR EXPRESSION
10 matches (9.86ms)
` ^(\S+) (\S+) (\S+) \[([^\]]+)\] "([^"]*)" (\d+) (\d+)$ ` gm
TEST STRING
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] "GET /shuttle/countdown/video/livevideo.gif HTTP/1.0" 200 0
205.212.115.106 - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
d104.aa.net - - [01/Jul/1995:00:00:13 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
129.94.144.152 - - [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
11:1
EXPLANATION
` ^(\S+) (\S+) (\S+) \[([^\]]+)\] ` gm
"([^"]*)" (\d+) (\d+)$
^ asserts position at start of a line
1st Capturing Group (\S+)
\S matches any non-whitespace character (equivalent to [^\r\n\t\f\v])
MATCH INFORMATION
Group 1 87-107 unicomp6.unicomp.net
Group 2 108-109 -
Group 3 110-111 -
Group 4 113-139 01/Jul/1995:00:00:06 -0400
QUICK REFERENCE
Search reference
All Tokens
Common Tokens
General Tokens
Anchors
A single ch... [abc]
A charact... [^abc]
A character... [a-z]
A charact... [^a-z]
A chara... [a-zA-Z]

CODE

parser/parser.go


THIS IS FINE.

CODE

loader/simple

CODE

cmd/load

```
$ du -sh logs
2.2Glogs
```


BIG
DATA

CODE

loader/seq

Load logs from file so we can **query** them.

CODE

loader/mapper

cel.dev

Overview Learn GitHub Contact us Search / English SIGN IN

Fast, safe expression language

Common Expression Language (CEL) is an expression language that's fast, portable, and safe to execute in performance-critical applications. CEL is designed to be embedded in an application, with application-specific extensions, and is ideal for extending declarative configurations that your applications might already use.

Use CEL for things like list filters for API calls, validation constraints on protocol buffers, and authorization rules for API requests.

LEARN MORE GET STARTED

```
// Simple predicates
'tacocat'.startsWith('taco')

// Parameterized predicates over structured data
account.balance >= transaction.withdrawal

// JSON objects
{'sub': '12345678',
 'aud': 'example2.cel.dev',
 'iss': 'https://example1.cel.dev/jwt-issuer'}

// Strongly typed objects
common.GeoPoint{ latitude: 10.0, longitude: -5.5 }
```

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.

Is CEL right for your project?

CEL is ideal for performance-critical applications because it was designed to evaluate safely and quickly (nanoseconds to

github.com/google/cel-go

Transfer-Encoding: chunked

jsonlines

ndjson

CODE

cmd/httpd

Questions?

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