

Regular Expressions (RegEx)

Regular expression is like another language

- What is a regular expression?
- Literal (or normal characters)
 - Alphanumeric
 - abc...ABC...0123...
 - Punctuation
 - `-_.,:;=()/+ *%&{}[]? ! $ ' ^ | \ < > " @ #`
- Just like languages Regular expressions also have dialects
 - awk, egrep, Emacs, grep, Perl, POSIX, Tcl, PROSITE

Why use a regular expression?

ALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP
EPTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETGKEIL
FYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETLAEY
KLKKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEM
LIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYVDYY
CRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNC
AADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAKPGV
WPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFY
PHGLGHGMGIDCHEIAGWPRGTCRGKKPHHSFVRFGRLE
KGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTV
GGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKESKL

| | | | | | | |
|------|------------|------|-------|--------------|-----------|---|
| FGCV | WKAQLLNEY | VAVK | FPIQ | DKQSWQNEY | EVYSLPGM | K |
| YGEV | WRGSWQGEN | VAVK | FSSR | DEKSWFRET | ELYNTVML | R |
| FGKV | YRAFWIGDE | VAVK | ARHDP | DEDISQTIENVR | QEAKLFAML | K |
| FGTV | YKKGKWHGD | VAVK | LKVV | DPTPEQFQAFRN | EVAVLRKT | R |
| SGTV | YKGVLEDDRH | VAVK | LENV | RQGKEVFQA | ELSVIGRI | N |

VAVK

Why use a regular expression?

ALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP
EPTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETGKEIL
FYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETLAEY
KLKKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEM
LIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYVDYY
CRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNC
AADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAKPGV
WPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFY
PHGLGHGMGIDCHEIAGWPRGTCRGKKPHHSFVRFGRLE
KGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTV
GGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKESKL

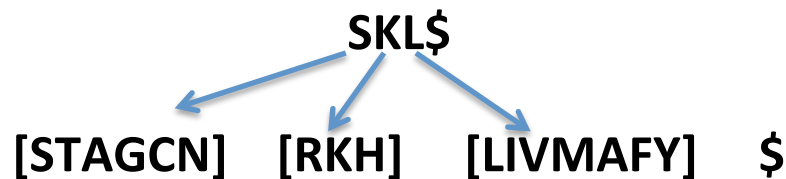
Why use a regular expression?

ALDVANRPMPKPEMFAAHRAKT LAELRKRKLEGVVLIYGFP
EPTRAHCD FEPVFRQESCFYWLTGVNEAD CAYFLDIETGKEIL
FYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETLAEY
KLKKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEM V
LIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYVDYY
CRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNC
AADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAKPGV
WPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFY
PHGLGHGMGIDCHEIAGWPRGTCRGKKPHHSFVRFGR TLE
KGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTV
GGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKE**SKL**

SKL\$

Why use a regular expression?

ALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP
EPTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETGKEIL
FYDPDIPQAYIIWFGELATIDDIKKKYGFEDVRLMPKIQETLAEY
KLKKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEM
LIQYACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYVDYY
CRCFAFSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNC
AADNTRTIPANGKFSPDQRAVYQAVLDCHNYVVAHAKPGV
WPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFY
PHGLGHGMGIDCHEIAGWPRGTCRGKKPHHSFVRFGRTE
KGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTV
GGVRIEDDLLITEDGCKVLSNIPKEIDEIEAFMAKKE**SKL**



- **AL**STDNVANRPMPKPEMF....
- Text: The sequence must start with an alanine, followed by any amino acid, followed by a serine or a threonine, two times, followed by any amino acid or nothing, followed by any amino acid except a valine.
- Regex: **A**.**[ST]**{2}.?**[^V]**