

# 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?

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRAHCDFEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVEANLIKHYVDYYCRCFA  
FSTIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTR  
TIPANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGV  
WPDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPH  
GLGHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRTE  
KGVVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTVGG  
VRIEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# Why use a regular expression?

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKFEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFY  
PDIPQAYIIWFGELATIDDI QQQQQ GFEDVRLMPKIQETLAEYK  
LKKIHTLPETCILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQ  
YACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAF  
STIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTI  
PANGKFSP QQQQQ RAVYQAVVAVKLDCHNYVVAHAKPGVW  
PDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGL  
GHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRTLEKG  
VVITNEPGCYFIRPSYNAAFADPEKSKYINKEVCERLRKTVGGVR  
IEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# Why use a regular expression?

## *To find a pattern*

MALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP  
EPTRDRINKEPVFRQESCFYWLTGVNEADCAFLDIETGKEILFY  
PDIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYK  
LKKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQ  
YACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAF  
STIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTI  
PANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVW  
PDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGL  
GHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRTLEKG  
VVITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGV  
RIEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# Why use a regular expression?

## *To find a pattern*

MALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP  
EPTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFY  
PDIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYK  
LKKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQ  
YACDVNSFAVRDTFKKVHPKMWEHQVMILKHYVDYYCRCFAF  
STIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTI  
PANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVW  
PDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGL  
GHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRTLEKG  
VVITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGV  
RIEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

FGCV	WKAQLLNEY	VAVK	FPIQ	DKQSWQNEY	EVYS	LPGM	K
YGEV	WRGSWQGEN	VAVK	FSSR	DEKSWFRET	ELYNTVML	R	
FGKV	YRAFWIGDE	VAVK	ARHDP	DEDISQTIENVR	Q	EAKLFAML	K
FGTV	YK GKWHGD	VAVK	LKVV	DPTPEQFQAFRN	EVAVLRKT	R	
SGTV	YKGVLEDDRH	VAVK	LENV	RQGKEVFQA	ELSVIGRI	N	

VAVK

# Why use a regular expression?

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGY VAVK DKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAV VAVK LDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL



# Why use a regular expression?

MALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKDEL

***KDEL\$***

***[HK]DEL\$***

- **MLSTDNVANRPMKPEMF....**
- **Text:** The sequence must start with an methionine, 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:** `^M.[ST]{2}.?[^V]`