

# ***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?*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCTRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Why use a regular expression?*

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCTRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Why use a regular expression?*

## *To find a pattern*

MALDVANRPMPKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCTRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Why use a regular expression?*

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP  
EPTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFY  
PDIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYK  
LKKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQ  
YACDVNSFAVRDTFKKVHPKMWEHQVMILKHYVDYYCRCFAF  
STIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTI  
PANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVW  
PDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGL  
GHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRRTLEKG  
VVITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGV  
RIEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Why use a regular expression?*

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFP  
EPTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFY  
PDIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYK  
LKKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQ  
YACDVNSFAVRDTFKKVHPKMWEHQVMILKHYVDYYCRCFAF  
STIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTI  
PANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVW  
PDLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGL  
GHGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRRTLEKG  
VVITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGV  
RIEDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Where do patterns come from?*

FGCV	WKAQLLNEY	VAVKIFPIQ	DKQSWQNEY	EVYSLPGM	K
YGEV	WRGSWQGEN	VAVKIFSSR	DEKSWFRET	ELYNTVML	R
FGKV	YRAFWIGDE	VAVKAARHDP	DEDISQTIENV	QEAKLFAML	K
FGTV	YKGGKWHGD	VAVKILKVV	DPTPEQFQAFRN	EVAVLRKT	R
SGTV	YKGVLEDDRH	VAVKLENV	RQGKEVFQA	ELSVIGRI	N



# *Where do patterns come from?*



# *Why use a regular expression?*

## *To find a pattern*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCRGKKPHHSFVRFGRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKESKL

# *Why use a regular expression?*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHVVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCTRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAKKDEL

***KDEL\$***

# *Why use a regular expression?*

MALDVANRPMKPEMFAAHRAKTLAELRKRKLEGVVLIYGFPE  
PTRDRINKEPVFRQESCFYWLTGVNEADCAYFLDIETGKEILFYP  
DIPQAYIIWFGELATIDDIQQQQQGFEDVRLMPKIQETLAEYKL  
KKIHTLRKRKILKGYVAVKDKNEFIDVVGELRQIKDDDEMVLIQY  
ACDVNSFAVRDTFKKVHPKMWEHQVMILKHYVDYYCRCFAFS  
TIVCSGENCSILHYHHNNKFIEDGELILIDTGCEYNCAADNTRTIP  
ANGKFSPQQQQQRAVYQAVVAVKLDCHNYVVAHAKPGVWP  
DLAYDSAKVMAAGLLKLGLFQNGTVDEIVDAGALAVFYPHGLG  
HGMGIDCHEIAHRAKGWPRGTCTRGKKPHHSFVRFGRRTLEKGV  
VITNEPGCYFIRPSYNAAFADPEKSKYRKRKVCERLRKTVGGVRI  
EDDLLITEDGCKVLSNIPKEIHRAKDEIEAFMAK**KDEL**

***KDEL\$***

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

- **ML**STDNVANRPMKPEMF....
- 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]**