



System Management Fundamentals

使用**Unix Shell**整理系统日志
Use Unix Shell to wrangle log data

May 2021

Microsoft Reactor | Ryan Chung

```
led by player to  
s.load_image("kg.png")  
  
[self]:  
ialize Dog object and create Text o  
g, self).__init__(image = Dog.image  
x = games.mouse.x  
bottom = games.sc  
  
re = games.Text(value = 0, size = 24  
top = 5, right = gam  
  
reen.add(self.score)  
1 = games.Text(value = 0, size = 24  
top = 5, left = gam
```



Ryan Chung

Instructor / DevelopIntelligence
Founder / MobileDev.TW

@ryanchung403 on WeChat
Ryan@MobileDev.TW





Reactor



developer.microsoft.com/reactor/
@MSFTReactor on Twitter



使用 UNIX shell 整理日志数据

2 分钟 剩余 • 模块 • 已完成 7 个单元, 共 9 个

★★★★☆ 4.5 (12)

对其进行评级

初级

学生

数据科学家

数据工程师

开发人员

Visual Studio Code

了解如何使用 UNIX shell 整理数据和检查数据文件。

学习目标

在本模块中, 你将学习如何:

- 使用命令 (如 `head`、`tail`、`wc`、`less` 和 `sort`) 执行基本文件检查。
- 使用 `cat` 创建、追加、显示和连接文件。
- 编写和使用 `regex` (正则表达式) 进行文本模式匹配。
- 使用 `grep` 搜索文件或使用 `stdin` 和 `regex` 进行模式匹配。
- 使用 `sed` 和 `regex` 对输入流执行基本文本转换。
- 使用管道和筛选器进行数据整理。

先决条件

- 命令行和基本 shell 命令的基本知识

学习目标

- 使用指令进行档案观察
- 建立、新增、显示、串接档案
- 正规表达式 (Regular Expression)
- 档案搜寻
- 文字转换
- 管线、筛选器

操作环境 - Bash

- Azure Cloud Shell
 - <https://shell.azure.com>
 - 或至 MS Learn Sandbox使用

<https://docs.microsoft.com/en-us/learn/modules/unix-shell-wrangle-data/1-file-inspection>

- Git BASH
- Cmder

确认操作环境

- 输入 `echo $SHELL`

```
/usr/bin/bash
```

取得数据

- 建立data资料夹

`mkdir data`

- 下载数据集

`wget -P data/ https://raw.githubusercontent.com/MicrosoftDocs/mslearn-data-wrangling-shell/main/NASA-logs-1995.txt`

`wget -P data/ https://raw.githubusercontent.com/MicrosoftDocs/mslearn-data-wrangling-shell/main/NASA-software-API.txt`

- 确认已下载

`cd data`

`ls`

```
NASA-logs-1995.txt  NASA-software-API.txt
```

两个档案

- NASA-Software-API.txt
 - NASA使用软体清单
- NASA-logs-1995.txt
 - NASA Kennedy 太空中心伺服器请求日志纪录

UNIX系统中的三个数据流

- 输入

`stdin` – 标准输入(键盘)

- 输出

`stdout` – 终端机输出

`stderr` – 状态回报，亦透过终端机输出

头尾瞧一瞧

- Head

head -n 5 NASA-software-API.txt

```
ARC-14136-1 ARC 2001-10-19T00:00:00.000 "Academic Worldwide" "Adaptive Relevance-Learning Software Component (ARNIE)"
ARC-14293-1 ARC 2005-09-19T00:00:00.000 "Open Source" "Genetic Graphs (JavaGenes)"
ARC-14297-1 ARC 2003-11-06T00:00:00.000 "General US" "Automated Domain Decomposition Software, PEGASUS Version 5.0"
ARC-14379-1 ARC 2002-03-27T00:00:00.000 "General US" "Man-machine Integration Design And Analysis System (MIDAS)"
ARC-14400-1 ARC 2001-01-29T00:00:00.000 "General US" "PLOT3D Version 4.0"
```

- tail

tail -n 5 NASA-software-API.txt

```
SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to SupportIn-Flight Characterization of Remote Sensing Systems"
SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"
GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"
GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics Data Facility Web Services"
GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```

一行一行来

- nl (numbers the lines)

nl -s = NASA-software-API.txt

```
695=SSC-00339 SSC 2009-10-05T00:00:00.000 "General Public" "A Tool for the Automated Verification of Spatial Resolution in Remotely Sensed Imagery"  
696=SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to Support In-Flight Characterization of Remote Sensing Systems"  
697=SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"  
698=GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"  
699=GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics DataFacility Web Services"  
700=GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```

行数、字数、字节数统计

- WC

wc NASA-software-API.txt

```
703 8917 81115 NASA-software-API.txt
```

行数 字数 字节数

行数、字数、字节数统计

- WC

wc NASA-software-API.txt

```
703 8917 81115 NASA-software-API.txt
```

行数 字数 字节数

等等，刚刚最后一笔不是第700行吗？怎么有703行

一行一行来并且要包含空行

- nl

nl -b a -s = NASA-software-API.txt

```
698=SSC-00339 SSC 2009-10-05T00:00:00.000 "General Public" "A Tool for the Automated Verification of Spatial Resolution in Remotely Sensed Imagery"  
699=SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to Support In-Flight Characterization of Remote Sensing Systems"  
700=SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"  
701=GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"  
702=GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics DataFacility Web Services"  
703=GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```


用cat检视档案、建立档案、新增内容至档案

- 检视档案内容

cat NASA-software-API.txt

```
SSC-00156-1 SSC 2003-04-24T00:00:00.000 "General Public" Storeplex
SSC-00161-1 SSC 2003-05-23T00:00:00.000 "General Public" "Generalized Computer Based Computation Of Venturi And Orifice Pressure Drops"
SSC-00181 SSC 2007-06-18T00:00:00.000 "General Public" "Application Research Toolbox (ART)"
SSC-00339 SSC 2009-10-05T00:00:00.000 "General Public" "A Tool for the Automated Verification of Spatial Resolution in Remotely Sensed Imagery"
SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to SupportIn-Flight Characterization of Remote Sensing Systems"
SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"
GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"
GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics Data Facility Web Services"
GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```

用cat检视档案、建立档案、新增内容至档案

- 建立档案
 - `cat > file1`
 - 按下Enter进入下一行后输入：`NASA headquarters`
 - 再按下Enter，然后按下Ctrl + D 离开

```
@Azure:~/data$ cat > file1
NASA headquarters
@Azure:~/data$ cat file1
NASA headquarters
```

用cat检视档案、建立档案、新增内容至档案

- 新增内容至档案
 - `cat >> file1`
 - 按下Enter进行下一行后输入：`Goddard Space Flight Center`
 - 再按下Enter，然后按下Ctrl + D 离开

```
@Azure:~/data$ cat >> file1
Goddard Space Flight Center
@Azure:~/data$ cat file1
NASA headquarters
Goddard Space Flight Center
```

用cat串接档案

- 串接两个档案的内容
 - `cat file1 file1`

```
@Azure:~/data$ cat file1 file1
NASA headquarters
Goddard Space Flight Center
NASA headquarters
Goddard Space Flight Center
```

用cat串接档案

- 串接两个档案的内容，并且输出成另一档案
- `cat file1 file1 > file2`

```
@Azure:~/data$ cat file1 file1 > file2
@Azure:~/data$ cat file2
NASA headquarters
Goddard Space Flight Center
NASA headquarters
Goddard Space Flight Center
```

正规表达式 Regular Expression

• 测试语法的好地方

The screenshot shows the regex101.com interface. The top navigation bar includes links for @regex101, donate, sponsor, contact, bug reports & feedback, wiki, and whats new?. The main interface is divided into several sections:

- SAVE & SHARE**: Includes a 'Save Regex' button with a keyboard shortcut 'ctrl+s'.
- FLAVOR**: A list of programming languages and versions supported, including PCRE2 (PHP >=7.3), PCRE (PHP <7.3), ECMAScript (JavaScript), Python 2.7, Golang, and Java 8.
- FUNCTION**: A list of regex functions, with 'Match' selected and marked with a green checkmark.
- TOOLS**: A section for additional tools, including a 'SPONSOR' button and the 'MOOVWEB' logo with the tagline 'Jamstack at Scale'.
- REGULAR EXPRESSION**: The main input area where the regex `^a.*` is entered. It shows '2 matches, 11 steps (~0ms)'.
- TEST STRING**: A text area containing the string `abc\nbcd\naaa`, with the matches highlighted in blue.
- EXPLANATION**: A section providing a detailed breakdown of the regex components:
 - `^` asserts position at start of a line.
 - `a` matches the character 'a' literally (case sensitive).
 - `.*` matches any character (except for line terminators) zero or more times.
- MATCH INFORMATION**: A table showing the matches found:

Match	Index	Text
Match 1	0-3	abc
Match 2	8-11	aaa
- QUICK REFERENCE**: A section for quick access to common regex tokens and their syntax.

<https://regex101.com/>

开启 code 检视 NASA-software-API.txt 档案

- 开启档案
 - code NASA-software-API.txt

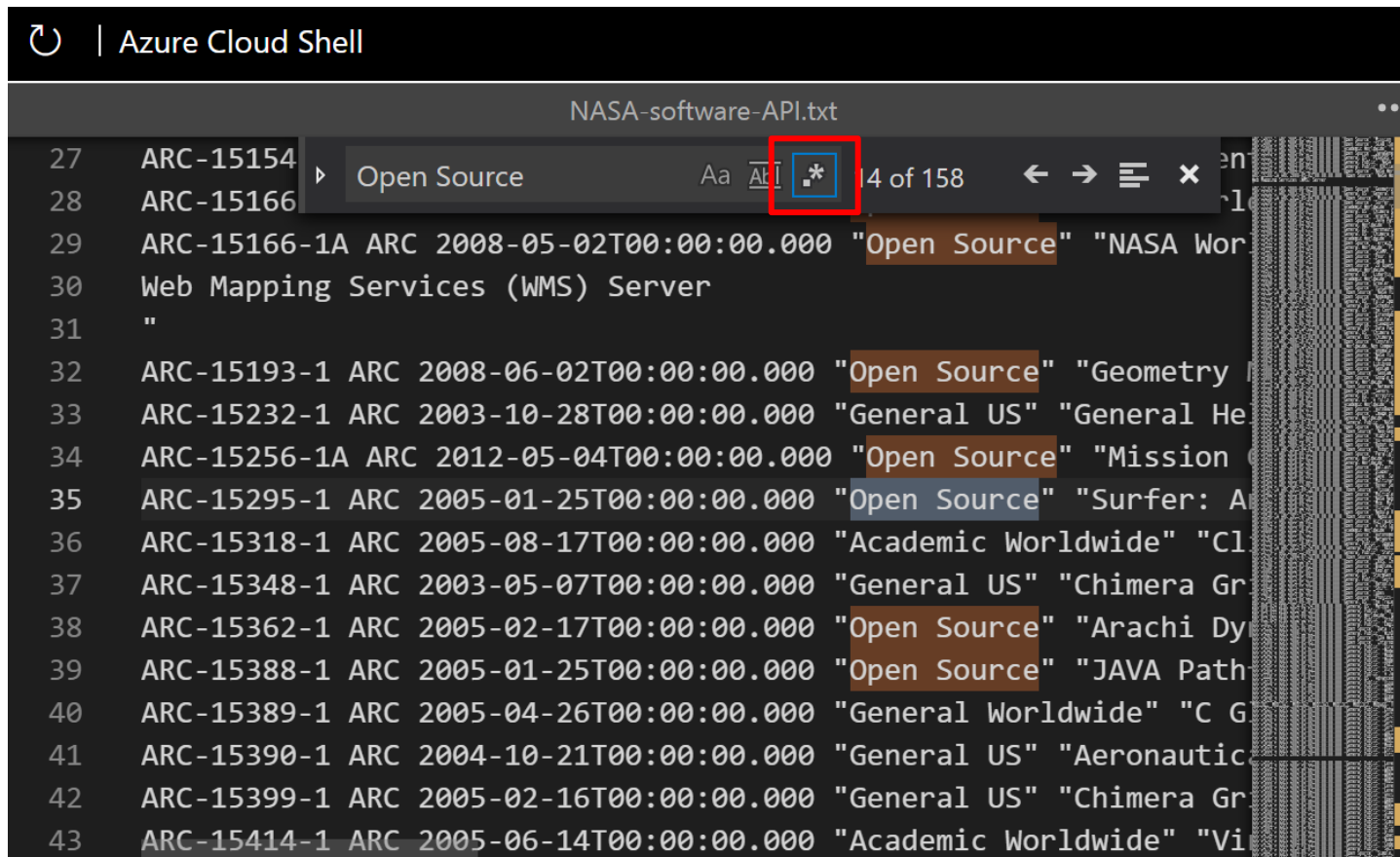
```
🔄 | Azure Cloud Shell

NASA-software-API.txt

1  ARC-14136-1 ARC 2001-10-19T00:00:00.000 "Academic Worldwide" "Ad
2  ARC-14293-1 ARC 2005-09-19T00:00:00.000 "Open Source" "Genetic G
3  ARC-14297-1 ARC 2003-11-06T00:00:00.000 "General US" "Automated
4  ARC-14379-1 ARC 2002-03-27T00:00:00.000 "General US" "Man-machin
5  ARC-14400-1 ARC 2001-01-29T00:00:00.000 "General US" "PLOT3D Ver
6  ARC-14467-1A ARC 2014-01-24T00:00:00.000 "Open Source" "CFD Util
7  ARC-14487-1 ARC 2005-02-14T00:00:00.000 "Open Source" "CAPTools-
8  "
9  ARC-14490-1 ARC 2001-11-13T00:00:00.000 "Academic Worldwide" "Li
10 ARC-14529-1 ARC 2004-01-21T00:00:00.000 "Open Source" "IND 2.1 -
11 ARC-14583-1 ARC 2002-06-17T00:00:00.000 "General Worldwide" "SLA
12 ARC-14598-1 ARC 2002-05-30T00:00:00.000 "General Worldwide" "Spa
13 ARC-14649-1 ARC 2001-12-17T00:00:00.000 "General US" "ILab Param
14 ARC-14653-1 ARC 2004-06-18T00:00:00.000 "Academic Worldwide" "Fu
15 ARC-14654-1 ARC 2001-09-13T00:00:00.000 "Academic Worldwide" "DI
16 ARC-14672-1 ARC 2003-12-15T00:00:00.000 "Open Source" "A Softwar
17 ARC-14725-1 ARC 2004-01-21T00:00:00.000 "Open Source" "Livingsto
```

搜寻关键字，并确认RegEx功能开启

- Ctrl+F or Command+F，搜寻Open Source
- 并确认RegEx功能有开启



Azure Cloud Shell

NASA-software-API.txt

27 ARC-15154

28 ARC-15166

29 ARC-15166-1A ARC 2008-05-02T00:00:00.000 "Open Source" "NASA Wor

30 Web Mapping Services (WMS) Server

31 "

32 ARC-15193-1 ARC 2008-06-02T00:00:00.000 "Open Source" "Geometry

33 ARC-15232-1 ARC 2003-10-28T00:00:00.000 "General US" "General He

34 ARC-15256-1A ARC 2012-05-04T00:00:00.000 "Open Source" "Mission

35 ARC-15295-1 ARC 2005-01-25T00:00:00.000 "Open Source" "Surfer: A

36 ARC-15318-1 ARC 2005-08-17T00:00:00.000 "Academic Worldwide" "Cl

37 ARC-15348-1 ARC 2003-05-07T00:00:00.000 "General US" "Chimera Gr

38 ARC-15362-1 ARC 2005-02-17T00:00:00.000 "Open Source" "Arachi Dy

39 ARC-15388-1 ARC 2005-01-25T00:00:00.000 "Open Source" "JAVA Path

40 ARC-15389-1 ARC 2005-04-26T00:00:00.000 "General Worldwide" "C G

41 ARC-15390-1 ARC 2004-10-21T00:00:00.000 "General US" "Aeronautic

42 ARC-15399-1 ARC 2005-02-16T00:00:00.000 "General US" "Chimera Gr

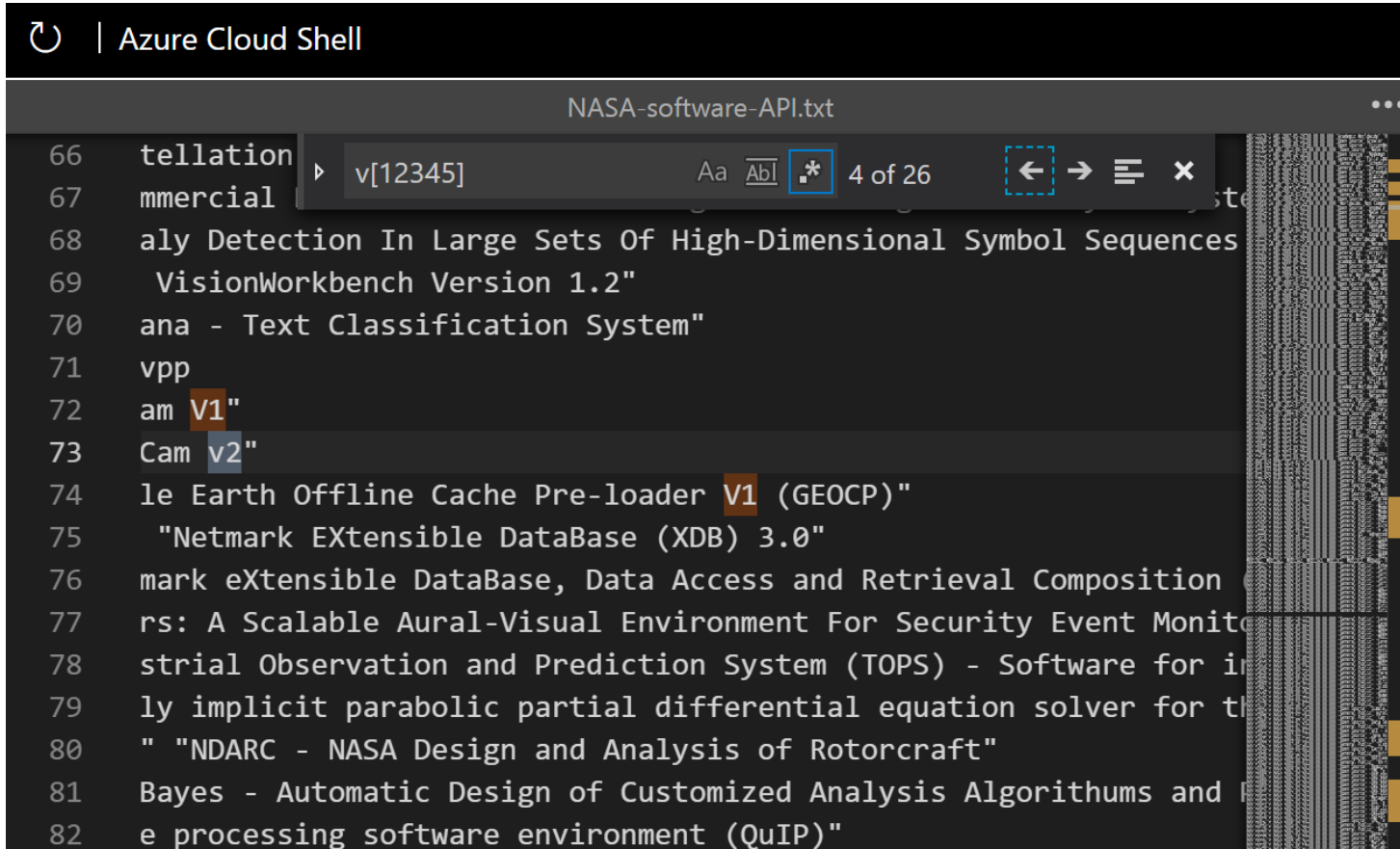
43 ARC-15414-1 ARC 2005-06-14T00:00:00.000 "Academic Worldwide" "Vi

查找版本号

- 想要找v1, v2, v3, v4, v5

- 搜寻 v[12345]

[] 任何在这list中的字元



```

Azure Cloud Shell

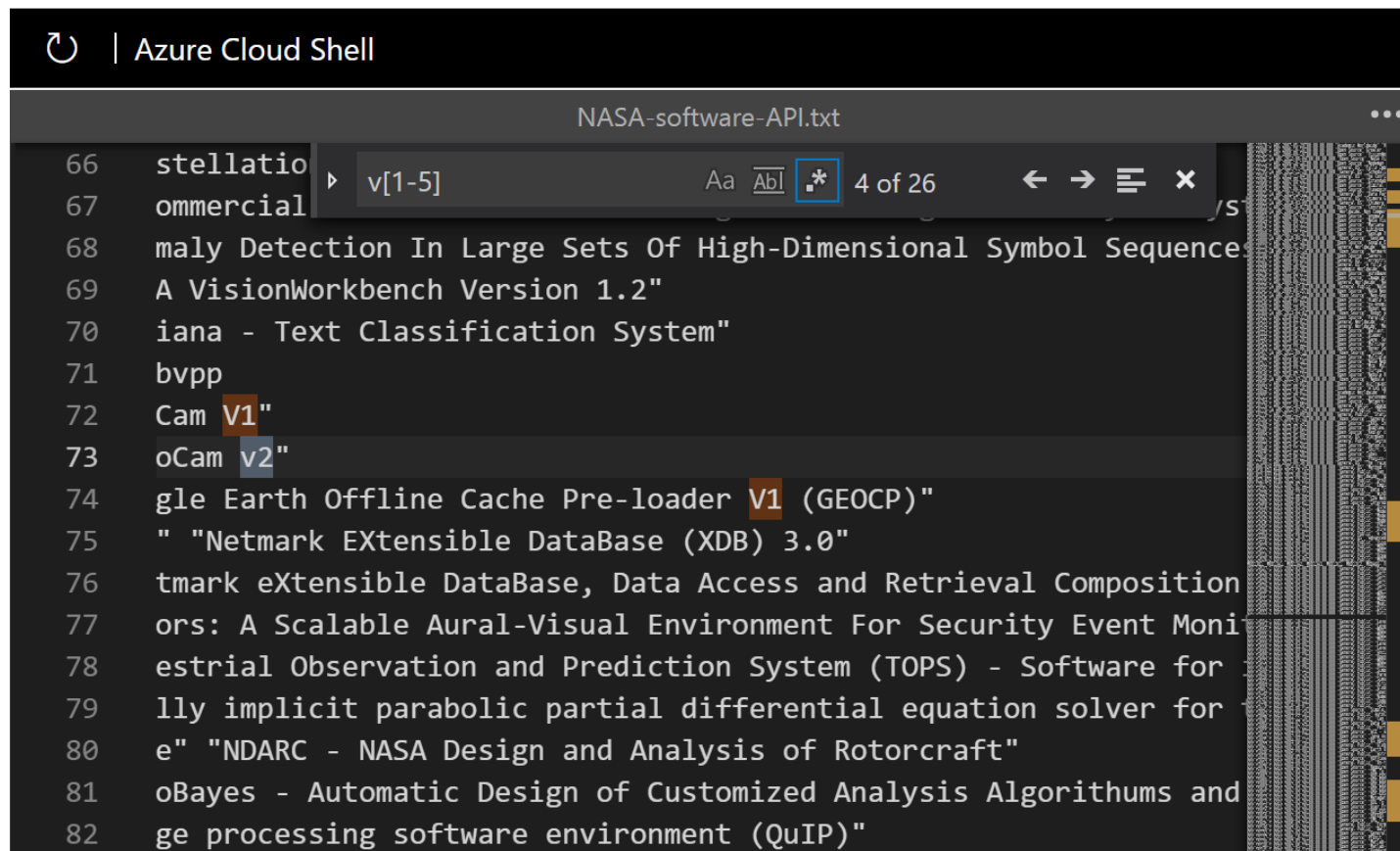
NASA-software-API.txt
66 tellation v[12345]
67 mmercial
68 aly Detection In Large Sets Of High-Dimensional Symbol Sequences
69 VisionWorkbench Version 1.2"
70 ana - Text Classification System"
71 vpp
72 am V1"
73 Cam V2"
74 le Earth Offline Cache Pre-loader V1 (GEOCP)"
75 "Netmark eXtensible DataBase (XDB) 3.0"
76 mark eXtensible DataBase, Data Access and Retrieval Composition
77 rs: A Scalable Aural-Visual Environment For Security Event Monito
78 strial Observation and Prediction System (TOPS) - Software for in
79 ly implicit parabolic partial differential equation solver for th
80 " "NDARC - NASA Design and Analysis of Rotorcraft"
81 Bayes - Automatic Design of Customized Analysis Algorithms and
82 e processing software environment (QuIP)"

```

查找版本号

- 想要找v1, v2, v3, v4, v5
- 也可以写成 v[1-5]

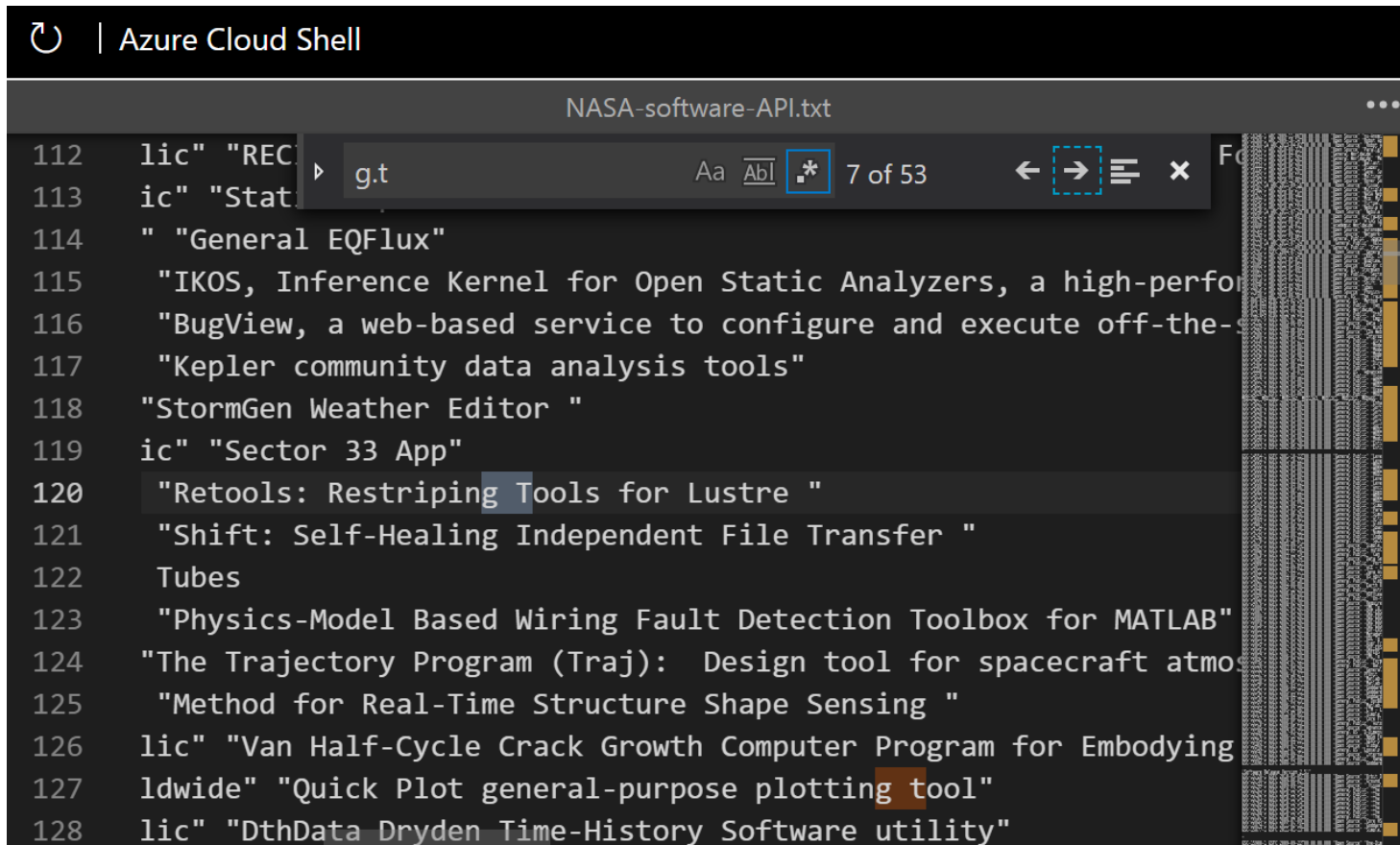
[] 任何在这list中的字元



```
Azure Cloud Shell
NASA-software-API.txt
66 stellatio
67 ommercial
68 maly Detection In Large Sets Of High-Dimensional Symbol Sequence
69 A VisionWorkbench Version 1.2"
70 iana - Text Classification System"
71 bvpp
72 Cam V1"
73 oCam v2"
74 gle Earth Offline Cache Pre-loader V1 (GEOCP)"
75 " "Netmark EXtensible DataBase (XDB) 3.0"
76 tmark eXtensible DataBase, Data Access and Retrieval Composition
77 ors: A Scalable Aural-Visual Environment For Security Event Moni
78 estrial Observation and Prediction System (TOPS) - Software for
79 lly implicit parabolic partial differential equation solver for
80 e" "NDARC - NASA Design and Analysis of Rotorcraft"
81 oBayes - Automatic Design of Customized Analysis Algorithms and
82 ge processing software environment (QuIP)"
```

. wildcard (任意字元/数字/空白/换行/标点/符号)

- 我想要找g开头、中间1个任意字元、t结尾的
- 搜寻 g.t



Azure Cloud Shell

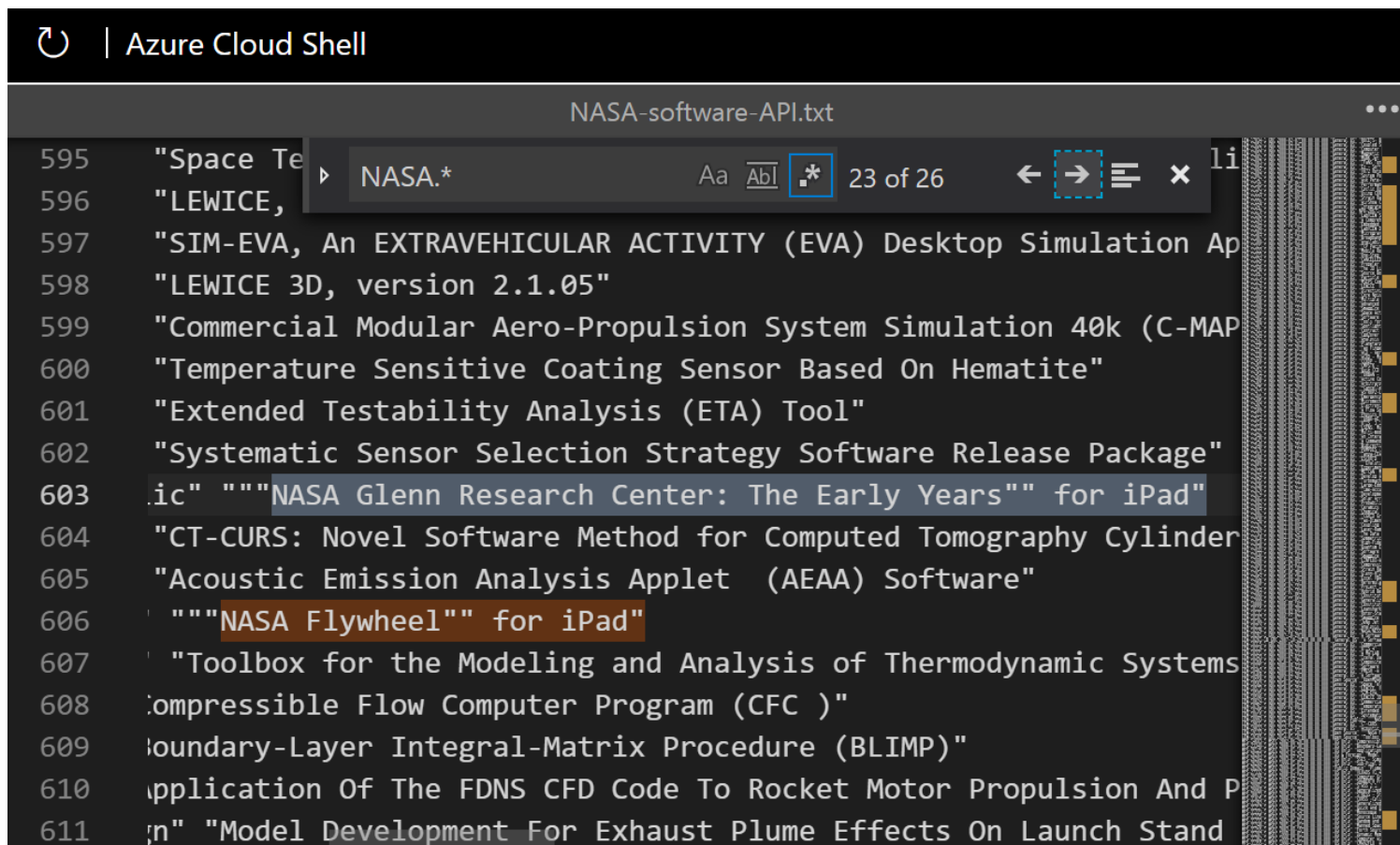
NASA-software-API.txt

g.t 7 of 53

```
112 lic" "REC
113 ic" "Stat
114 " "General EQFlux"
115 "IKOS, Inference Kernel for Open Static Analyzers, a high-perfor
116 "BugView, a web-based service to configure and execute off-the-s
117 "Kepler community data analysis tools"
118 "StormGen Weather Editor "
119 ic" "Sector 33 App"
120 "Retools: Restripping Tools for Lustre "
121 "Shift: Self-Healing Independent File Transfer "
122 Tubes
123 "Physics-Model Based Wiring Fault Detection Toolbox for MATLAB"
124 "The Trajectory Program (Traj): Design tool for spacecraft atmos
125 "Method for Real-Time Structure Shape Sensing "
126 lic" "Van Half-Cycle Crack Growth Computer Program for Embodying
127 ldwide" "Quick Plot general-purpose plotting tool"
128 lic" "DthData Dryden Time-History Software utility"
```


* 0 ~ 多次

- 我想要找NASA后面跟着0个或多个任意字元均可
- 搜寻 NASA.*



```
595 "Space Te
596 "LEWICE,
597 "SIM-EVA, An EXTRAVEHICULAR ACTIVITY (EVA) Desktop Simulation Ap
598 "LEWICE 3D, version 2.1.05"
599 "Commercial Modular Aero-Propulsion System Simulation 40k (C-MAP
600 "Temperature Sensitive Coating Sensor Based On Hematite"
601 "Extended Testability Analysis (ETA) Tool"
602 "Systematic Sensor Selection Strategy Software Release Package"
603 ic" ""NASA Glenn Research Center: The Early Years"" for iPad"
604 "CT-CURS: Novel Software Method for Computed Tomography Cylinder
605 "Acoustic Emission Analysis Applet (AEAA) Software"
606 ""NASA Flywheel"" for iPad"
607 "Toolbox for the Modeling and Analysis of Thermodynamic Systems
608 Compressible Flow Computer Program (CFC )"
609 boundary-Layer Integral-Matrix Procedure (BLIMP)"
610 Application Of The FDNS CFD Code To Rocket Motor Propulsion And P
611 n" "Model Development For Exhaust Plume Effects On Launch Stand
```


^ 开头、\$结尾

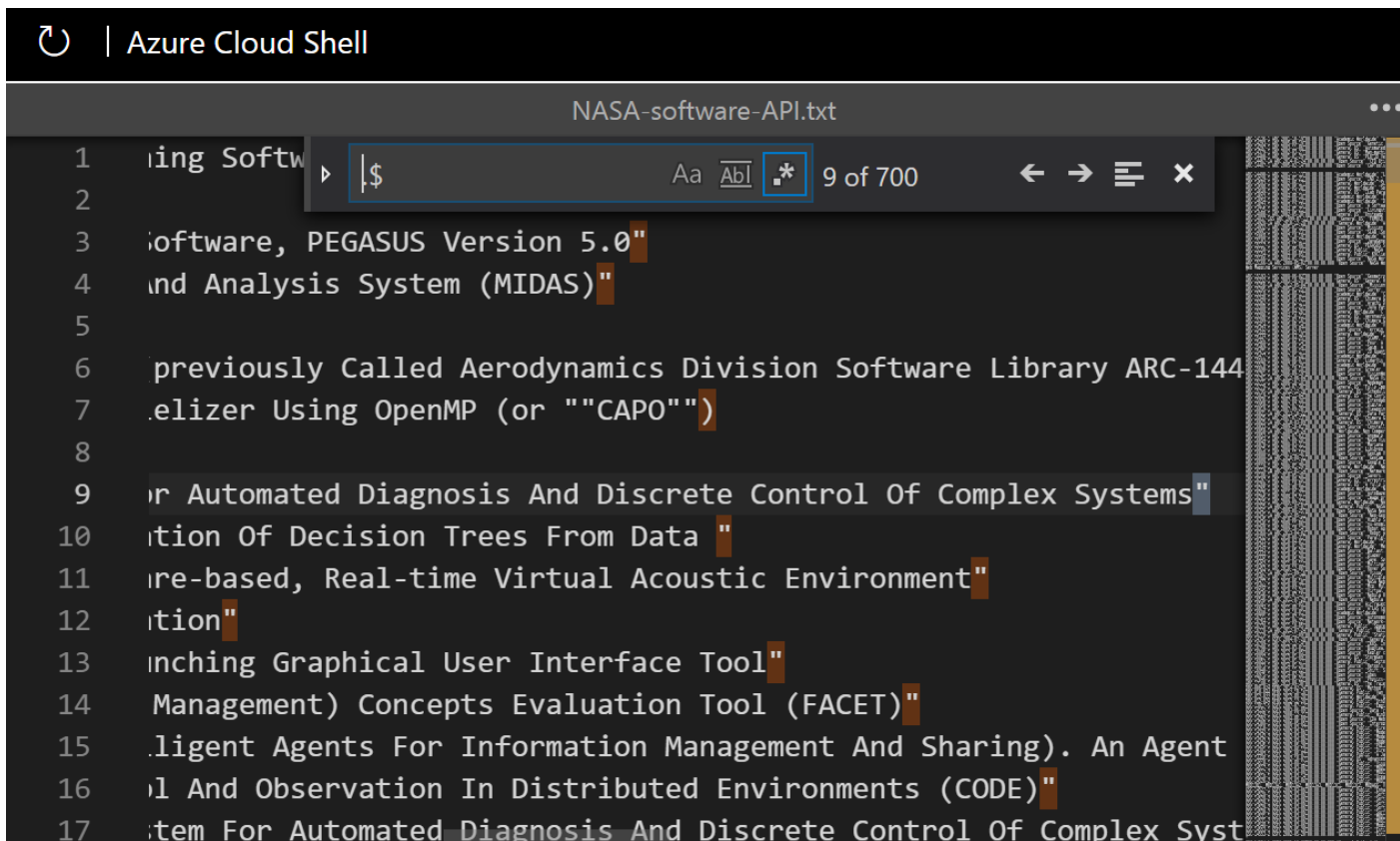
- 想要找0~9数字开头的
 - 搜寻 `^[0-9]`
- 想要找0~9数字结尾的
 - 搜寻 `[0-9]$`

```
Azure Cloud Shell
NASA-software-API.txt
258 GSC-16228 ^[0-9]
259 GSC-16230
260 GSC-16232-1 GSFC 2011-10-18T00:00:00.000 "Open Source" "The Core
261 GSC-16321-1 GSFC 2012-02-21T00:00:00.000 "General Public" "Space
262 GSC-16565-1 GSFC 2012-05-21T00:00:00.000 "Open Source" "General
263 GSC-16612-1 GSFC 2012-10-10T00:00:00.000 "Open Source" "Space Op
264 GSC-16668-1 GSFC 2013-01-15T00:00:00.000 "General Public" "The Na
265 produced by the NASA GSFC Storytelling Team. This application is
266 2011."
267 GSC-16683-1 GSFC 2013-05-30T00:00:00.000 "Open Source" "An Andro
268 "
269 GSC-16824-1 GSFC 2013-06-20T00:00:00.000 "Open Source" "Evolution
270 GSC-16848-1 GSFC 2013-07-29T00:00:00.000 "Open Source" "Multi-mi
271 HQN-11316-1 HDQS 2009-09-08T00:00:00.000 "General Worldwide" "NA
272 KSC-11777 KSC 2006-10-31T00:00:00.000 "General Public" SpecsInta
273 KSC-12043 KSC 2000-12-13T00:00:00.000 "General US" "Ground Proce
274 KSC-12079 KSC 2003-08-26T00:00:00.000 "General US" "Personnel Lo
```

```
Azure Cloud Shell
NASA-software-API.txt
221
222 [0-9]$
223
224
225 arison software"
226
227
228 Flight Safety System"
229 cation Programming Interface (GMSEC Architecture API) [R3] 2.5
230
231
232
233 ire"
234 (ViSBARD) "
235 ickMap"
236
237 IM"
```

\ 跳脱字元

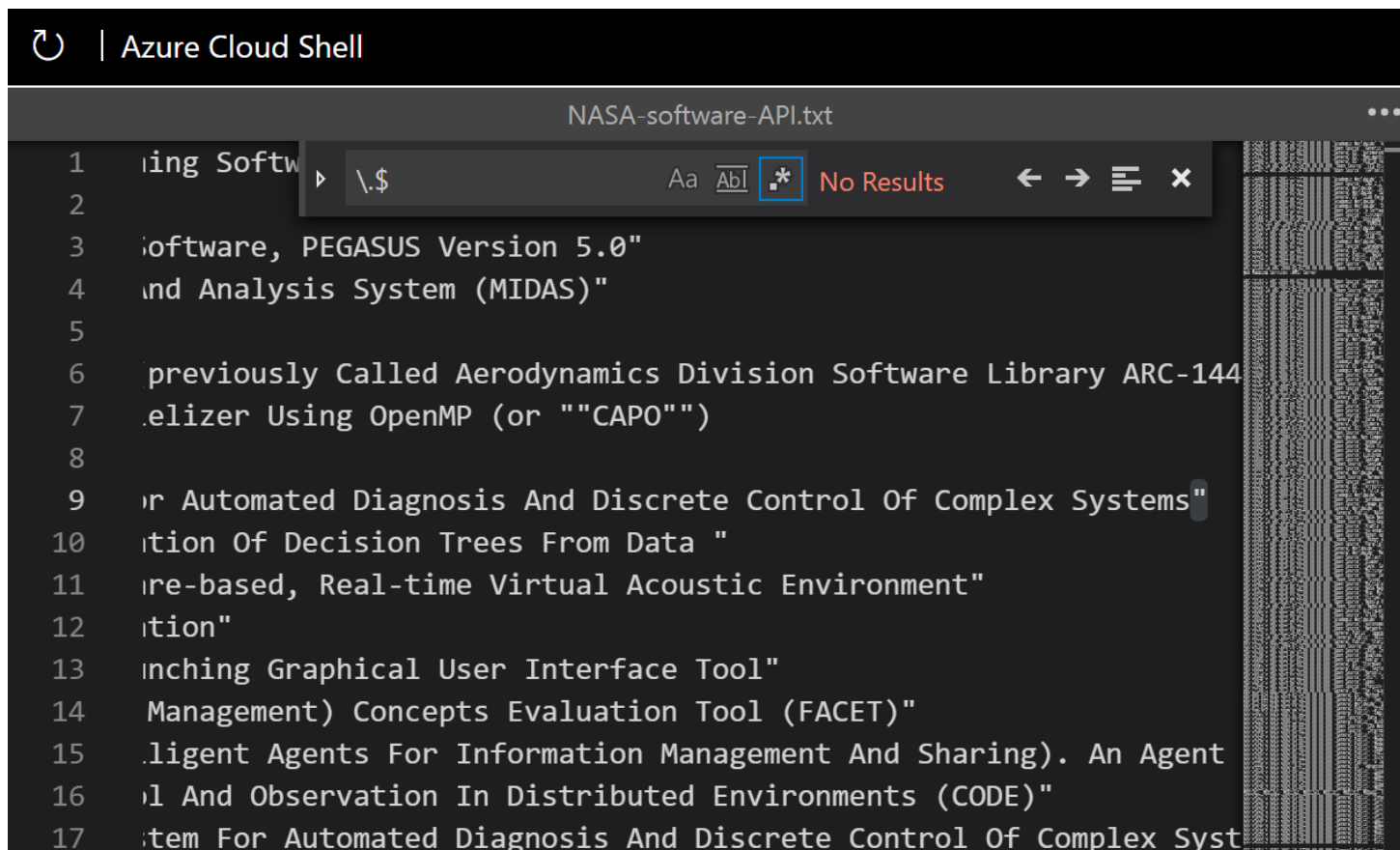
- 我想要找 . 结尾的
- 搜寻 .\$ 结果全部都来了(任意字元结尾)



```
Azure Cloud Shell
NASA-software-API.txt
1  ing Softw
2  |$
3  Software, PEGASUS Version 5.0"
4  and Analysis System (MIDAS)"
5
6  previously Called Aerodynamics Division Software Library ARC-144
7  elizer Using OpenMP (or ""CAPO"")
8
9  or Automated Diagnosis And Discrete Control Of Complex Systems"
10 tion Of Decision Trees From Data "
11 re-based, Real-time Virtual Acoustic Environment"
12 tion"
13 nching Graphical User Interface Tool"
14 Management) Concepts Evaluation Tool (FACET)"
15 lligent Agents For Information Management And Sharing). An Agent
16 l And Observation In Distributed Environments (CODE)"
17 tem For Automated Diagnosis And Discrete Control Of Complex Syst
```

\ 跳脱字元

- 我想要找 . 结尾的
- 搜寻 \.\$ 确保是要搜寻 . 结尾的



Azure Cloud Shell

NASA-software-API.txt

```
1  ing Softw
2
3  software, PEGASUS Version 5.0"
4  and Analysis System (MIDAS)"
5
6  previously Called Aerodynamics Division Software Library ARC-144
7  elizer Using OpenMP (or ""CAPO"")
8
9  or Automated Diagnosis And Discrete Control Of Complex Systems"
10 tion Of Decision Trees From Data "
11 re-based, Real-time Virtual Acoustic Environment"
12 tion"
13 nching Graphical User Interface Tool"
14 Management) Concepts Evaluation Tool (FACET)"
15 lligent Agents For Information Management And Sharing). An Agent
16 l And Observation In Distributed Environments (CODE)"
17 stem For Automated Diagnosis And Discrete Control Of Complex Syst
```

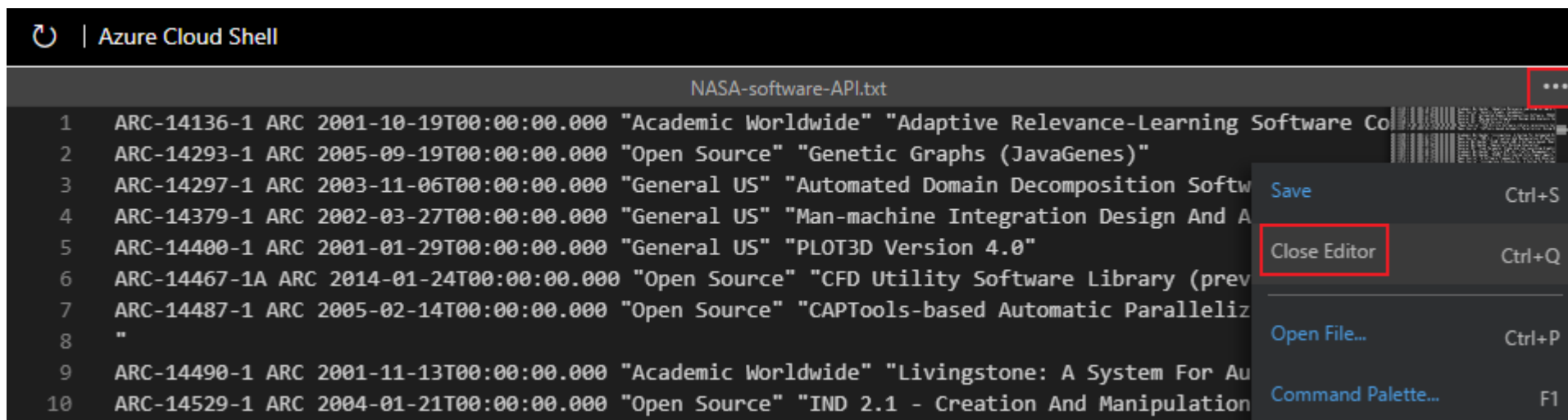
Search: \.\$ No Results

RegEx Cheat Sheet

符号	意义	符号	意义
^	一行的开头	+	1 ~ 多次
\$	一行的结尾	+?	1 ~ 多次(non-greedy)
.	任意字元	[aeiou]	符合任一个在list中的字元
\s	空格	[^XYZ]	符合任一个不在list中的字元
\S	任何非空格字元	[a-z0-9]	表达范围
*	0 ~ 多次	(撷取字串的开头
*?	0 ~ 多次(non-greedy))	撷取字串的结尾

在Azure Cloud Shell中关闭Code

- Ctrl + q
- 点击右上方 . . .



The screenshot shows the Azure Cloud Shell interface. At the top, there's a header with a refresh icon and the text "Azure Cloud Shell". Below this, a file named "NASA-software-API.txt" is open in an editor. The editor displays a list of software entries, each with a line number (1-10), an ARC ID, a date, and a description. A context menu is open on the right side of the editor, triggered by clicking the three dots in the top right corner of the file's editor area. The menu options are: "Save" (Ctrl+S), "Close Editor" (Ctrl+Q), "Open File..." (Ctrl+P), and "Command Palette..." (F1). The "Close Editor" option is highlighted with a red box.

```
1  ARC-14136-1 ARC 2001-10-19T00:00:00.000 "Academic Worldwide" "Adaptive Relevance-Learning Software Co
2  ARC-14293-1 ARC 2005-09-19T00:00:00.000 "Open Source" "Genetic Graphs (JavaGenes)"
3  ARC-14297-1 ARC 2003-11-06T00:00:00.000 "General US" "Automated Domain Decomposition Softw
4  ARC-14379-1 ARC 2002-03-27T00:00:00.000 "General US" "Man-machine Integration Design And A
5  ARC-14400-1 ARC 2001-01-29T00:00:00.000 "General US" "PLOT3D Version 4.0"
6  ARC-14467-1A ARC 2014-01-24T00:00:00.000 "Open Source" "CFD Utility Software Library (prev
7  ARC-14487-1 ARC 2005-02-14T00:00:00.000 "Open Source" "CAPTools-based Automatic Paralleliz
8  "
9  ARC-14490-1 ARC 2001-11-13T00:00:00.000 "Academic Worldwide" "Livingstone: A System For Au
10 ARC-14529-1 ARC 2004-01-21T00:00:00.000 "Open Source" "IND 2.1 - Creation And Manipulation
```

使用 grep 查找档案内容

- 寻找 *.novo.dk

grep '*\.novo\.dk' NASA-logs-1995.txt

```
@Azure:~/data$ grep '\*\.novo\.dk' NASA-logs-1995.txt
***.novo.dk 805465029 GET /ksc.html 200 7067
***.novo.dk 805465031 GET /images/ksclogo-medium.gif 200 5866
***.novo.dk 805465051 GET /images/MOSAIC-logosmall.gif 200 363
***.novo.dk 805465053 GET /images/USA-logosmall.gif 200 234
***.novo.dk 805465054 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465058 GET /images/WORLD-logosmall.gif 200 669
***.novo.dk 805465068 GET /shuttle/missions/missions.html 200 8678
***.novo.dk 805465071 GET /images/launchmedium.gif 200 11853
***.novo.dk 805465153 GET /images/KSC-logosmall.gif 200 1204
***.novo.dk 805465157 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465323 GET /images/launch-logo.gif 200 1713
***.novo.dk 805465328 GET /history/apollo/images/apollo-logo1.gif 200 1173
***.novo.dk 805465381 GET /shuttle/resources/orbiters/columbia.html 2006922
***.novo.dk 807951768 GET /shuttle/missions/sts-69/mission-sts-69.html 200 11264
***.novo.dk 807951775 GET /shuttle/missions/sts-69/sts-69-patch-small.gif 200 8083
***.novo.dk 807951782 GET /images/KSC-logosmall.gif 200 1204
***.novo.dk 807951786 GET /images/launch-logo.gif 200 1713
***.novo.dk 807951792 GET /history/apollo/images/apollo-logo1.gif 200 1173
***.novo.dk 807951832 GET /shuttle/countdown/ 200 4673
***.novo.dk 807951848 GET /shuttle/missions/sts-69/count69.gif 200 46053
***.novo.dk 807951864 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 807951907 GET /htbin/cdt_main.pl 200 3714
***.novo.dk 807951914 GET /shuttle/countdown/images/countclock.gif 200 13994
***.novo.dk 807951938 GET /shuttle/countdown/liftoff.html 200 4665
***.novo.dk 807951981 GET /shuttle/countdown/video/livevideo2.gif 200 69067
***.novo.dk 807952044 GET /htbin/cdt_clock.pl 200 543
***.novo.dk 807952060 GET /shuttle/countdown/lps/fr.html 200 1879
***.novo.dk 807952078 GET /shuttle/countdown/lps/fr.gif 200 30232
***.novo.dk 807952102 GET /shuttle/countdown/lps/back.gif 200 1289
```


可同时运用于多个档案

- 同时在两个档案中找寻NASA，并显示符合的行号

```
grep -n 'NASA' NASA-logs-1995.txt NASA-software-API.txt
```

```
NASA-software-API.txt:499:LEW-17324-1 GRC 2001-01-05T00:00:00.000 "General US" "CANCELLED ** Same As LEW-16855-1 (APNASA - Average Passage Flow Solver)"
NASA-software-API.txt:571:LEW-17952-1 GRC 2005-08-18T00:00:00.000 "General US" "Excel VBA Add-in for Calculating the Drag Coefficient from WakeSurvey Experimental Data Taken at the NASA Glenn Research Center's Icing Research Tunnel"
NASA-software-API.txt:583:LEW-18229-1 GRC 2007-03-27T00:00:00.000 "General US" "NASA/NESSUS 6.2c Probabilistic Structural Analysis Software"
NASA-software-API.txt:603:LEW-18993-1 GRC 2012-09-12T00:00:00.000 "General Public" "" "NASA Glenn Research Center: The Early Years" for iPad"
NASA-software-API.txt:606:LEW-19133-1 GRC 2014-01-16T00:00:00.000 "OpenSource" "" "NASA Flywheel" for iPad"
NASA-software-API.txt:687:MSC-24585-1 JSC 2011-09-12T00:00:00.000 "General Public" "NASA Electronic Position Description System (ePDS)"
NASA-software-API.txt:693:MSC-25694-1 JSC 2013-09-10T00:00:00.000 "General Public" "NASA Hazard Management System (HMS)"
```

grep 参数

参数	意义
-c	计算行数
-n	显示行号与符合样式
-i	大小写均可
-w	整个字符符合
-H	包含档案名称
-m<max number lines>	显示行数上限，例如4行 -m4

sed 指令

- 输出档案内容

sed " NASA-software-API.txt

```
MSC-25184-1 JSC 2011-10-05T00:00:00.000 "General US" "Space Shuttle Cockpit Avionics Upgrade Display Atlas Disclosure (CAU Atlas)"
MSC-25185-1 JSC 2011-10-05T00:00:00.000 "General US" "FCOD Rapid Prototyping Lab Generic Display Software Disclosure (FCOD Generic displays)"
MSC-25186-1 JSC 2011-10-05T00:00:00.000 "General US" "eProc Electronic Procedure System for Spacecraft glass cockpits (eProc System)"
MSC-25694-1 JSC 2013-09-10T00:00:00.000 "General Public" "NASA Hazard Management System (HMS)"
SSC-00151-1 SSC 2003-04-23T00:00:00.000 "General Public" "Engineering Units Generator 'Eugen'"
SSC-00156-1 SSC 2003-04-24T00:00:00.000 "General Public" Storeplex
SSC-00161-1 SSC 2003-05-23T00:00:00.000 "General Public" "Generalized Computer Based Computation Of Venturi And Orifice Pressure Drops"
SSC-00181 SSC 2007-06-18T00:00:00.000 "General Public" "Application Research Toolbox (ART)"
SSC-00339 SSC 2009-10-05T00:00:00.000 "General Public" "A Tool for the Automated Verification of Spatial Resolution in Remotely Sensed Imagery"
SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to Support In-Flight Characterization of Remote Sensing Systems"
SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"
GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"
GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics Data Facility Web Services"
GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```

sed 指令

- 寻找内容并且替代

sed 's/Target/John/'

```
@Azure:~$ echo "My name is Ryan" | sed 's/Ryan/John/'  
My name is John
```

- 练习

sed 's/NASA/National Aerospace Agency/' NASA-software-API.txt

sed 指令

- 寻找内容并且替代

`sed 's/Target/John/'`

```
@Azure:~$ echo "My name is Ryan" | sed 's/Ryan/John/'  
My name is John
```

- 练习

`sed 's/NASA/National Aerospace Agency/' NASA-software-API.txt`

- 练习二：只印出有替代的那几行

`sed -n 's/NASA/National Aerospace Agency/p' NASA-software-API.txt`

sed 指令

- 寻找内容并且替代

`sed 's/Target/John/'`

```
@Azure:~$ echo "My name is Ryan" | sed 's/Ryan/John/'  
My name is John
```

- 练习

`sed 's/NASA/National Aerospace Agency/' NASA-software-API.txt`

- 练习二：只印出有替代的那几行

`sed -n 's/NASA/National Aerospace Agency/p' NASA-software-API.txt`

- 练习三：替代并且输出至档案

`sed -n 's/NASA/National Aerospace Agency/w NASA-replaced.txt' NASA-software-API.txt`

管线练习

- 将NASA-software-API.txt的内容送去加上行号

```
cat NASA-software-API.txt | nl
```

```
695 SSC-00339 SSC 2009-10-05T00:00:00.000 "General Public" "A Tool for the Automated Verification of Spatial Resolution in Remotely Sensed Imagery"  
696 SSC-00393 SSC 2013-05-17T00:00:00.000 "General Public" "Software Suite to Support In-Flight Characterization of Remote Sensing Systems"  
697 SSC-00424 SSC 2013-09-06T00:00:00.000 "General Public" "SSC Site Status Mobile Application"  
698 GSC-14732-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Tool For Interactive Plotting, Sonification, And 3D Orbit Display (TIPSOD)"  
699 GSC-14730-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Space Physics Data Facility Web Services"  
700 GSC-14726-1 GSFC 2004-06-09T00:00:00.000 "Open Source" "Earth Observing System (EOS) Clearinghouse (ECHO)"
```

管线练习

- 将所有.txt档案计算行数，并从小排到大

```
wc -l *.txt | sort -n
```

```
@Azure:~/data$ wc -l *.txt | sort -n
  26 NASA-replaced.txt
 703 NASA-software-API.txt
200000 NASA-logs-1995.txt
200729 total
```

wc -l 计算行数(line)

sort -n 依数值(number)大小排序，预设为由小到大

sort -n -r 依数值(number)大小排序，由大到小

管线练习

- 将所有.txt档案计算行数，并从小排到大

```
wc -l *.txt | sort -n
```

```
@Azure:~/data$ wc -l *.txt | sort -n
  26 NASA-replaced.txt
 703 NASA-software-API.txt
200000 NASA-logs-1995.txt
200729 total
```

- 只显示前面2行

```
wc -l *.txt | sort -n | head -n 2
```

```
@Azure:~/data$ wc -l *.txt | sort -n | head -n 2
  26 NASA-replaced.txt
 703 NASA-software-API.txt
```

管线练习

- 将NASA-software-API.txt 每行加上行号与等号，然后只显示前5行

`nl -s = NASA-software-API.txt | head -n 5`

```
@Azure:~/data$ nl -s = NASA-software-API.txt | head -n 5
```

```
1=ARC-14136-1 ARC 2001-10-19T00:00:00.000 "Academic Worldwide" "Adaptive Relevance-Learning Software Component (ARNIE)"
2=ARC-14293-1 ARC 2005-09-19T00:00:00.000 "Open Source" "Genetic Graphs (JavaGenes)"
3=ARC-14297-1 ARC 2003-11-06T00:00:00.000 "General US" "Automated Domain Decomposition Software, PEGASUS Version 5.0"
4=ARC-14379-1 ARC 2002-03-27T00:00:00.000 "General US" "Man-machine Integration Design And Analysis System (MIDAS)"
5=ARC-14400-1 ARC 2001-01-29T00:00:00.000 "General US" "PLOT3D Version 4.0"
```

管线练习

- 将NASA-software-API.txt 每行加上行号与等号，然后只显示前5行，依行号从小排到大

```
nl -s = NASA-software-API.txt | head -n 5 | sort -n -r
```

```
@Azure:~/data$ nl -s = NASA-software-API.txt | head -n 5 | sort -n -r
```

```
5=ARC-14400-1 ARC 2001-01-29T00:00:00.000 "General US" "PLOT3D Version 4.0"
```

```
4=ARC-14379-1 ARC 2002-03-27T00:00:00.000 "General US" "Man-machine Integration Design And Analysis System (MIDAS)"
```

```
3=ARC-14297-1 ARC 2003-11-06T00:00:00.000 "General US" "Automated Domain Decomposition Software, PEGASUS Version 5.0"
```

```
2=ARC-14293-1 ARC 2005-09-19T00:00:00.000 "Open Source" "Genetic Graphs (JavaGenes)"
```

```
1=ARC-14136-1 ARC 2001-10-19T00:00:00.000 "Academic Worldwide" "Adaptive Relevance-Learning Software Component (ARNIE)"
```

管线练习

- 找出NASA-logs-1995.txt中有NASA的前10行

```
grep -m 10 'NASA' NASA-logs-1995.txt
```

```
@Azure:~/data$ grep -m 10 'NASA' NASA-logs-1995.txt
***.novo.dk 805465054 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465157 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 807951864 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 805065766 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 805072932 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 806532057 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 807996997 GET /images/NASA-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 805378672 GET /images/NASA-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 806878949 GET /images/NASA-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 809570230 GET /images/NASA-logosmall.gif 304 0
```

管线练习

- 找出NASA-logs-1995.txt中有NASA的前10行，
照数字排序

```
grep -m 10 'NASA' NASA-logs-1995.txt | sort -n
```

```
@Azure:~/data$ grep -m 10 'NASA' NASA-logs-1995.txt | sort -n
***.novo.dk 805465054 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465157 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 807951864 GET /images/NASA-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 805378672 GET /images/NASA-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 806878949 GET /images/NASA-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 809570230 GET /images/NASA-logosmall.gif 304 0
007.thegap.com 805065766 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 805072932 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 806532057 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 807996997 GET /images/NASA-logosmall.gif 200 786
```

管线练习

- 找出NASA-logs-1995.txt中有NASA的前10行，
照数字反向排序

```
grep -m 10 'NASA' NASA-logs-1995.txt | sort -n -r
```

```
@Azure:~/data$ grep -m 10 'NASA' NASA-logs-1995.txt | sort -n -r
007.thegap.com 807996997 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 806532057 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 805072932 GET /images/NASA-logosmall.gif 200 786
007.thegap.com 805065766 GET /images/NASA-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 809570230 GET /images/NASA-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 806878949 GET /images/NASA-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 805378672 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 807951864 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465157 GET /images/NASA-logosmall.gif 200 786
***.novo.dk 805465054 GET /images/NASA-logosmall.gif 200 786
```

管线练习

- 找出NASA-logs-1995.txt中有NASA的前10行，照数字反向排序，再把NASA都换成全名

```
grep -m 10 'NASA' NASA-logs-1995.txt | sort -n -r | sed -n 's/NASA/National Aerospace Agency/p'
```

```
@Azure:~/data$ grep -m 10 'NASA' NASA-logs-1995.txt | sort -n -r | sed -n 's/NASA/National Aerospace Agency/p'
007.thegap.com 807996997 GET /images/National Aerospace Agency-logosmall.gif 200 786
007.thegap.com 806532057 GET /images/National Aerospace Agency-logosmall.gif 200 786
007.thegap.com 805072932 GET /images/National Aerospace Agency-logosmall.gif 200 786
007.thegap.com 805065766 GET /images/National Aerospace Agency-logosmall.gif 200 786
01-dynamic-c.wokingham.luna.net 809570230 GET /images/National Aerospace Agency-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 806878949 GET /images/National Aerospace Agency-logosmall.gif 304 0
01-dynamic-c.wokingham.luna.net 805378672 GET /images/National Aerospace Agency-logosmall.gif 200 786
***.novo.dk 807951864 GET /images/National Aerospace Agency-logosmall.gif 200 786
***.novo.dk 805465157 GET /images/National Aerospace Agency-logosmall.gif 200 786
***.novo.dk 805465054 GET /images/National Aerospace Agency-logosmall.gif 200 786
```

Summary

- cat
 - 显示、串接、添加内容至档案
- pipe(|)
 - 管线输出变成另一指令的输入
- Regular Expression
 - 正规表达式
- grep(**G**lobal **R**egular **E**xpression **P**rint)
 - 查找档案资讯，常与管线搭配使用
- sed(stream editor)
 - 最常被用来做字串替代

学习运用以上指令与参数进行资讯筛选、查找

知识检查

1. 在当前目录中，我们想要查找包含最少行数的三个文件。 可使用此处列出的哪一个命令？

- ☐ `wc -l * > sort -n > head -n 3`
- ☐ `wc -l * | sort -n | head -n 1-3`
- ☐ `wc -l * | sort -n | head -n 3`
- ☐ `wc -l * | head -n 3 | sort -n`

2. 正则表达式 `Fr[ea]nc[eh]` 匹配的内容是什么？

- ☐ French, France, Frence, Franch
- ☐ Frenche, Franceh, Frenceh, Franche
- ☐ France, French
- ☐ Freanceh, Fraenche

知识检查

3. Grep 的 -v 选项用于反转模式匹配，以便仅输出不匹配模式的行。以下哪个命令可查找 /data 中其名称以 s.txt 结尾但不包含字符串 net 的所有文件？例如，shuttles.txt 或 software.txt，而不是 planets.txt。

- ☐ find data -name *s.txt | grep -v net
- ☐ grep -v 'net' \$(find data -name '*s.txt')
- ☐ find data -name '*s.txt' | grep -v net
- ☐ 以上说法都不正确

4. 假设要删除某些已处理的数据文件，并且只保留原始文件和处理脚本来保存存储。原始文件以 .dat 结尾，处理后的文件以 .txt 结束。以下哪种解决方案会删除所有已处理的数据文件，且仅删除已处理的数据文件？

- ☐ rm ?.txt
- ☐ rm *.txt
- ☐ rm *.txt
- ☐ rm .



Reactor



developer.microsoft.com/reactor/
[@MSFTReactor](#) on Twitter

议程结束 感谢聆听



请记得填写课程回馈问卷 (Event ID : **XXXXXX**)
<https://aka.ms/Reactor/Survey>

© 2019 Microsoft Corporation. All rights reserved. The text in this document is available under the Creative Commons Attribution 3.0 License, additional terms may apply. All other content contained in this document (including, without limitation, trademarks, logos, images, etc.) are not included within the Creative Commons license grant. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it. Some examples are for illustration only and are fictitious. No real association is intended or inferred. Microsoft makes no warranties, express or implied, with respect to the information provided here.