title: The Unix Shell
subtitle: Introducing the Shell
minutes: 10

The Unix Shell: Introducing the Shell

Learning Objectives

- Explain how the shell relates to the keyboard, the screen, the operating system, and users' programs.
- Explain when and why command-line interfaces should be used instead of graphical interfaces.

Rochelle Terman, a Recovering Political Scientist in Training, wants to do analysis of some New York Times she downloaded from LexisNexis (http://www.lexisnexis.com/hottopics/lnacademic/). LexisNexis only lets her download up to 500 articles at a time, but she wants to analyze several years of coverage. So she downloads several text files, each containing no more than 500 articles, along with metadata for each article (date, desk, author, length, subject, etc).

Each text file looks something like this:

1 of 500 DOCUMENTS

The New York Times
September 25, 2011 Sunday
Correction Appended
Late Edition - Final

Monitoring Rights In Chechen Region, A Month at a Time
BYLINE: By SETH MYDANS
SECTION: Section A; Column 0; Foreign Desk; Pg. 8
LENGTH: 1150 words

GROZNY, Russia -- They never go out alone, and when they are in their small apartment here in the capital of Chechnya, a flat screen on the wall displays a continuous feed from security cameras in the hall and stairway outside...

2 of 500 DOCUMENTS

The New York Times Blogs
(The Lede)
October 14, 2011 Friday

Human Rights Group Welcomes Obama's Decision to Send Troops to Uganda

BYLINE: ROBERT MACKEY

LENGTH: 465 words

As my colleagues Thom Shanker and Rick Gladstone report, "President Obama said Friday that he had ordered the deployment of 100 armed military advisers to ...

Now she wants to turn those big text files into a spreadsheet, each each row being an article, and columns for the text and metadata.

This requires her to:

1. Combine all the bulk downloads into one file.
2. Parse each field (BYLINE, LENTH, text, etc) into its own column.
3. Analyze, publish, get a great tenure track job, and retire early.

The problem is, Rochelle has procrastinated for the last year and now has to get this project done ASAP. If she completes steps 1-2 by hand, it will take her approximately a gazillian years.

While we won't be able to complete this project today, this lesson will be a step in the right direction. Today we will work mostly with the filesystem using the shell to move, copy, rename, and merge files. Although that may not sound super exciting, it contains the FUN!damentals of programming and teaches how to work in one of the most important environments for programming. The shell runs the other languages you may learn, such as Python, R, node.JS, etc., and can link them all together in a nice workflow, which can be automated, repeated, and scheduled. Additionally, many of you are probably interested in accessing remote servers on campus to run computationally-intensive scripts. All of this is done through the shell. Once you ssh into the server, you will not have a GUI with which to interact, but only the shell.

Computer-Human Interfaces

At a high level, computers do four things:

- run programs
- store data
- communicate with each other
- interact with us

They can do the last of these in many different ways, including direct brain-computer links and speech interfaces. Since these are still in their infancy,most of us use windows, icons, mice, and pointers. These technologies didn't become widespread until the 1980s, but their roots go back to Doug Engelbart's work in the 1960s, which you can see in what has been called "The Mother of All Demos (http://www.youtube.com/watch?v=a11JDLBXtPQ)".

Going back even further, the only way to interact with early computers was to rewire them. But in between, from the 1950s to the 1980s, most people used line printers. These devices only allowed input and output of the letters, numbers, and punctuation found on a standard keyboard, so programming languages and interfaces had to be designed around that constraint.

The Command Line

This kind of interface is called a **command-line interface**, or CLI,
to distinguish it from the **graphical user interface**, or GUI, that most people now use.

The heart of a CLI is a **read-evaluate-print loop**, or REPL: when the user types a command and then presses the enter (or return) key, the computer reads it, executes it, and prints its output.
The user then types another command,
and so on until the user logs off.

The Shell

This description makes it sound as though the user sends commands directly to the computer, and the computer sends output directly to the user. In fact,
there is usually a program in between called a **command shell**.

What the user types goes into the shell; it figures out what commands to run and orders the computer to execute them.

Note, the reason why the shell is called *the shell*: it encloses the operating system in order to hide some of its complexity and make it simpler to interact with.

A shell is a program like any other. What's special about it is that its job is to run other programs rather than to do calculations itself. The commands are themselves programs: when they terminate, the shell gives the user another prompt ($ on our systems).

Bash

The most popular Unix shell is **Bash**, the Bourne Again Shell (so-called because it's derived from a shell written by Stephen Bourne --- this is what passes for wit among programmers). Bash is the default shell on most modern implementations of **Unix**, and in most packages that provide Unix-like tools for Windows.

Why Use a Shell

Using Bash or any other shell sometimes feels more like programming than like using a mouse. Commands are terse (often only a couple of characters long),
their names are frequently cryptic, and their output is lines of text rather than something visual like a graph.

On the other hand, the shell allows us to combine existing tools in powerful ways with only a few keystrokes and to set up pipelines to handle large volumes of data automatically.

In addition, the command line is often the easiest way to interact with remote machines. As clusters and cloud computing become more popular for scientific data crunching, being able to drive them is becoming a necessary skill.

Adapted from: [Software Carpentry (http://software-carpentry.org/v5/novice/shell/00-intro.html)]