information technology & management

viabil iit y

INTRO TO OPEN SOURCE

ILLINOIS INSTITUTE OF TECHNOLOGY

Exploring Linux Filesystems

Sean Hughes-Durkin

ITMO/IT-O 456 Fall 2017

Information Technology & Management Programs

School of Applied Technology

Objectives

At the end of this lesson students should be able to:

- Describe / navigate the Linux directory structure using relative and absolute pathnames
- Describe the various types of Linux files
- View filenames and file types
- Use shell wildcards to specify multiple filenames

Objectives

At the end of this lesson students should be able to:

- Display the contents of text files and binary files
- Search text files for regular expressions using **`grep`**
- Use the **`vi`** editor to manipulate text files
- Identify common alternatives to the **`vi`** text editor used today

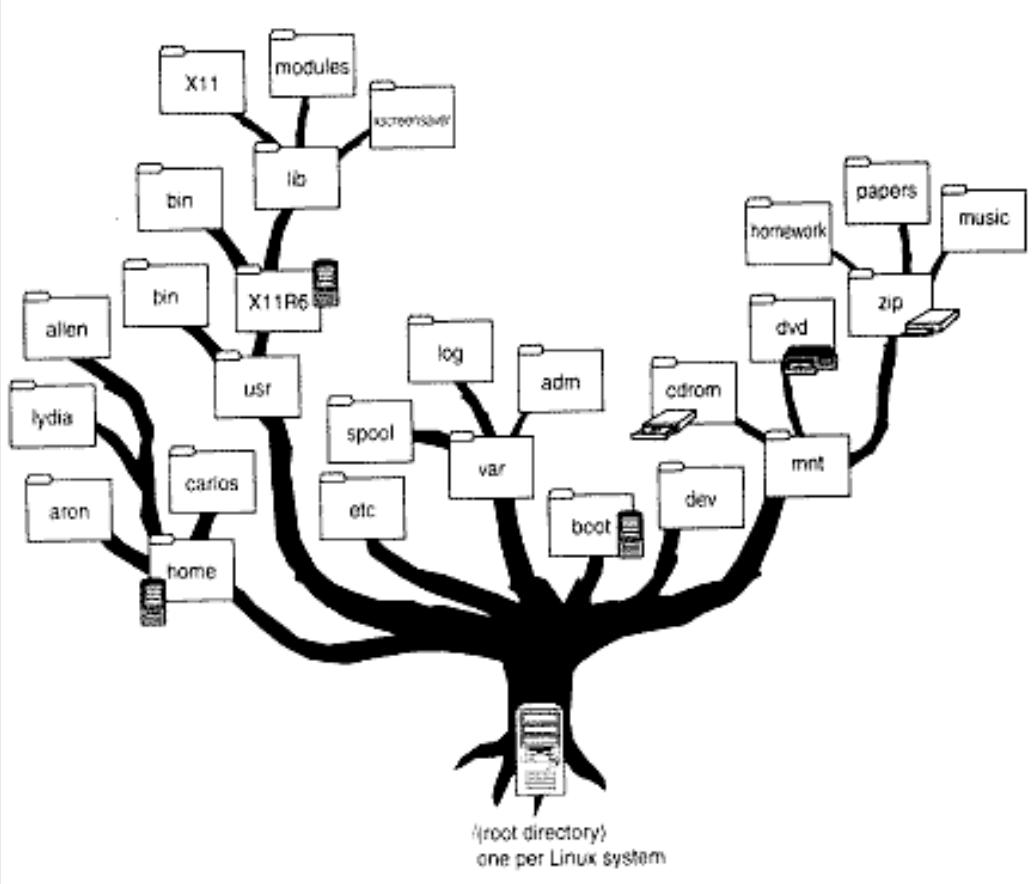
The Linux Directory Structure

- Directory
  - Used to organize other files or directories into a logical tree structure
  - Stored in a filesystem of a specific partition in the hard disk or SSD
  - Logical names applied for organization
    - Pictures, Videos, Music, etc

The Linux Directory Structure

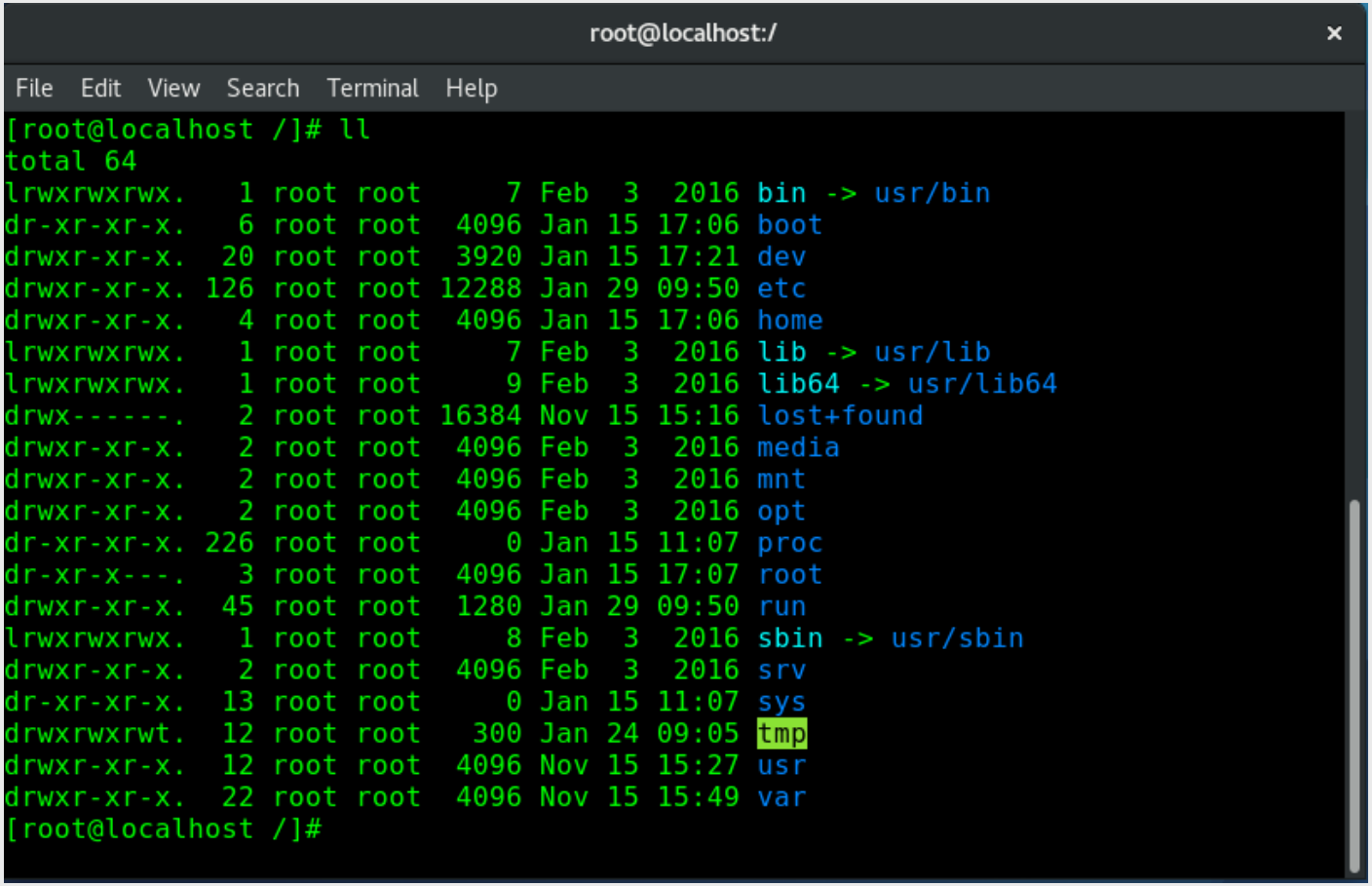
- Root of the filesystem
  - The top level directory
  - Referred to using the / character
  - Forms root of a hierarchical tree

The Linux Directory Structure



The Linux filesystem structure

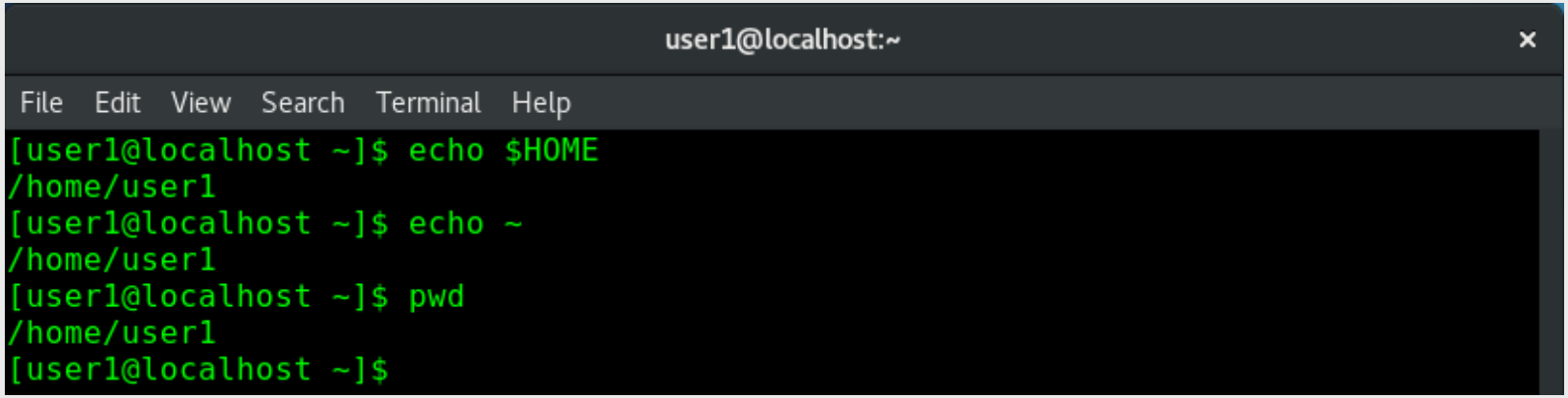
The Linux Directory Structure



Changing Directories

- Home directory
  - Unique to each user
  - **~** metacharacter used to refer to home directory
- **pwd** (print working directory) command
  - Displays current directory in the directory tree

Changing Directories



Changing Directories

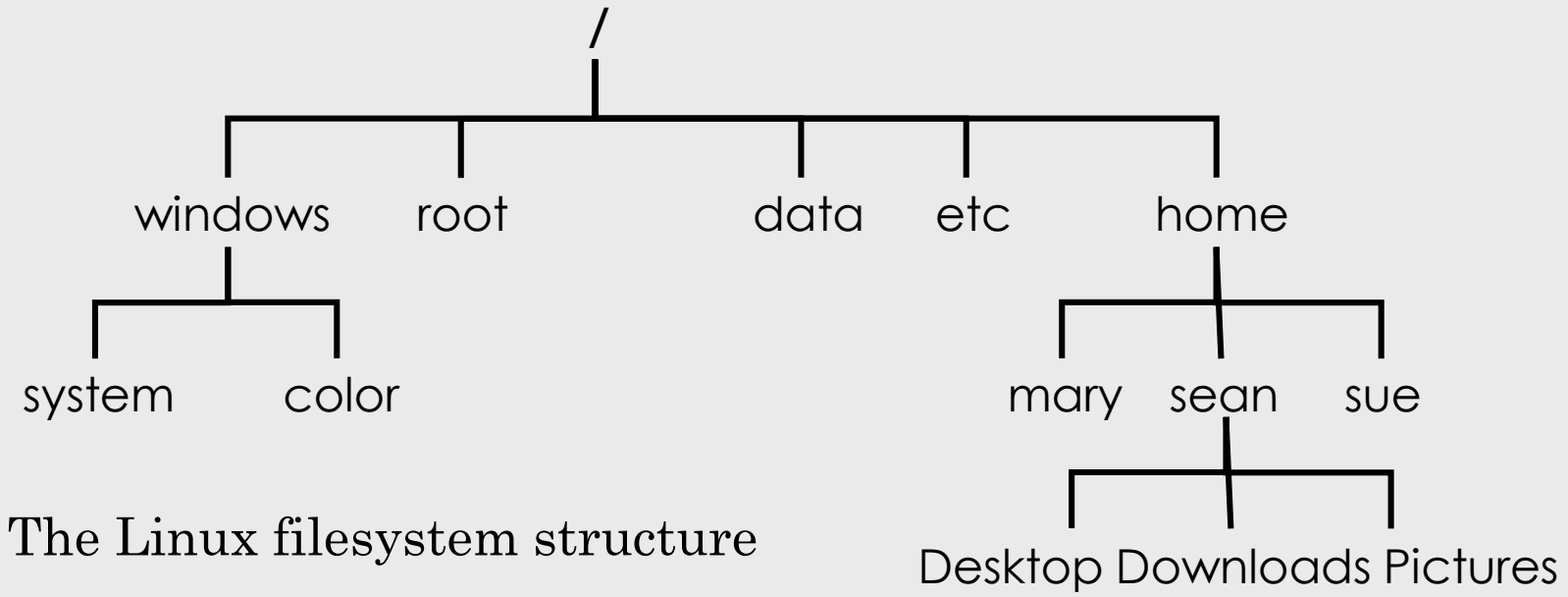
- **cd** (change directory) command
  - Change the current directory in the directory tree
  - Argument specifies the destination directory
    - Without an argument it will default to your home directory

Changing Directories

- Absolute pathname
  - Pathname from the root directory to a certain file or directory
  - Will always start with /
- Relative pathname
  - Pathname of file or directory relative to current directory
  - Sub-directories of the parent directory

The Linux Directory Structure

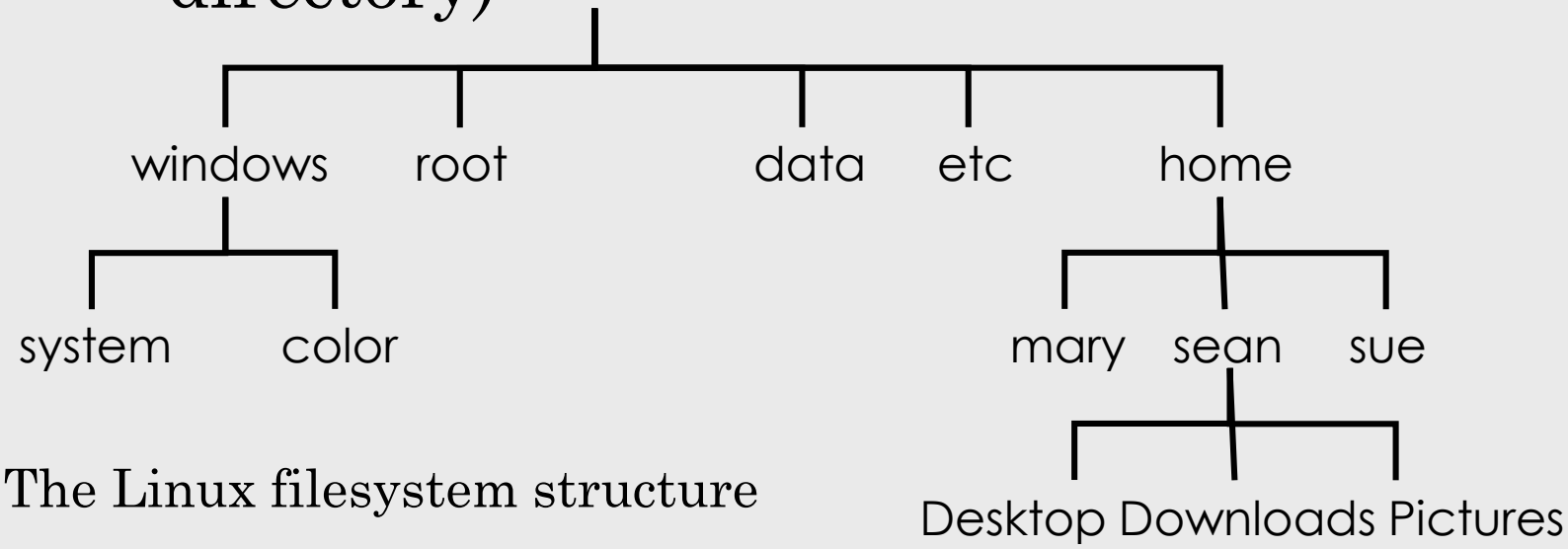
- Absolute pathname
  - Full pathname to a certain file or directory starting from the root directory
  - Example: **`/home/sean/Downloads`**



The Linux filesystem structure

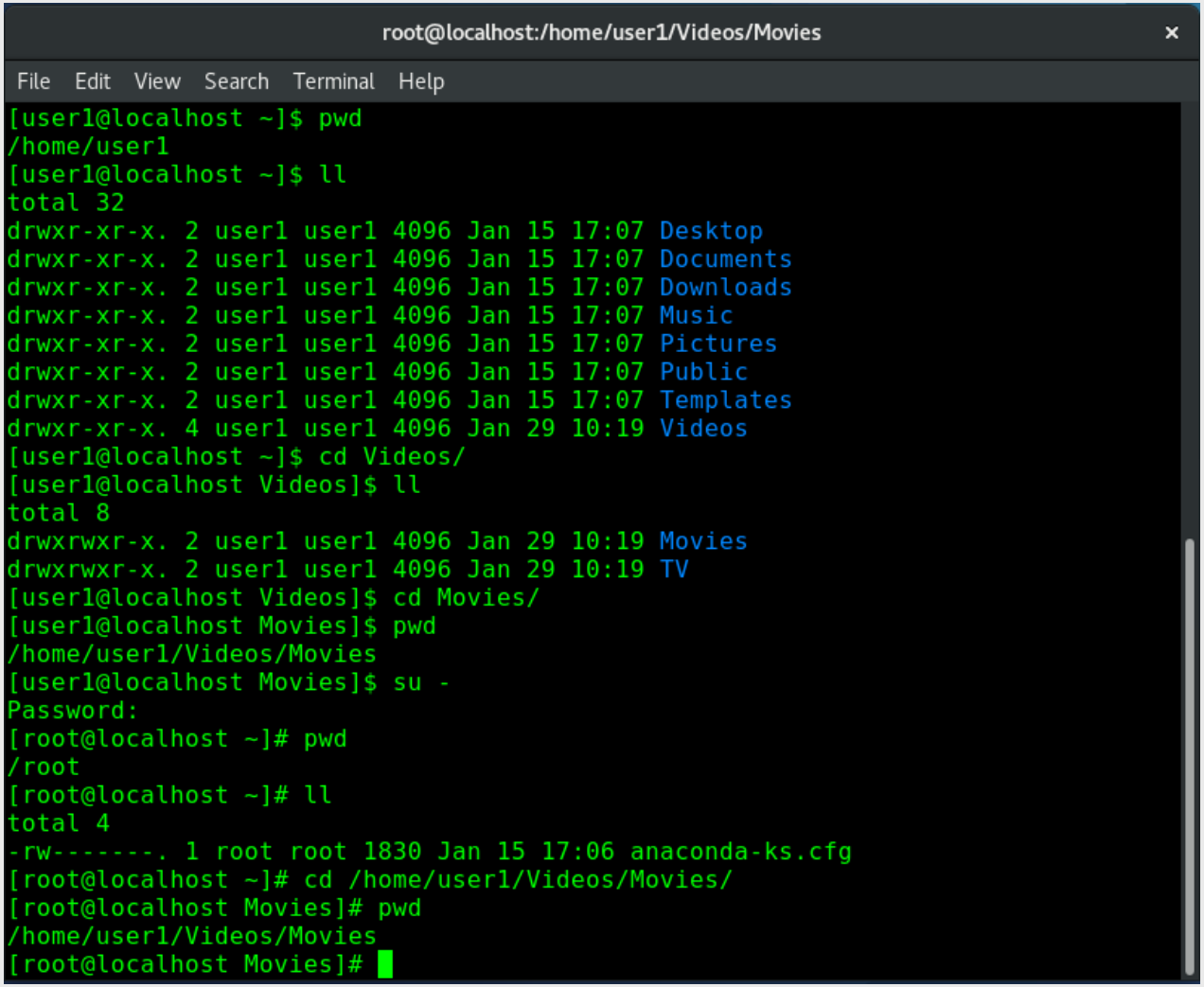
The Linux Directory Structure

- Relative pathname
  - Pathname to a certain file or directory starting from your current directory
  - Example: **`Downloads`** (from the **`sean`** directory)



The Linux filesystem structure

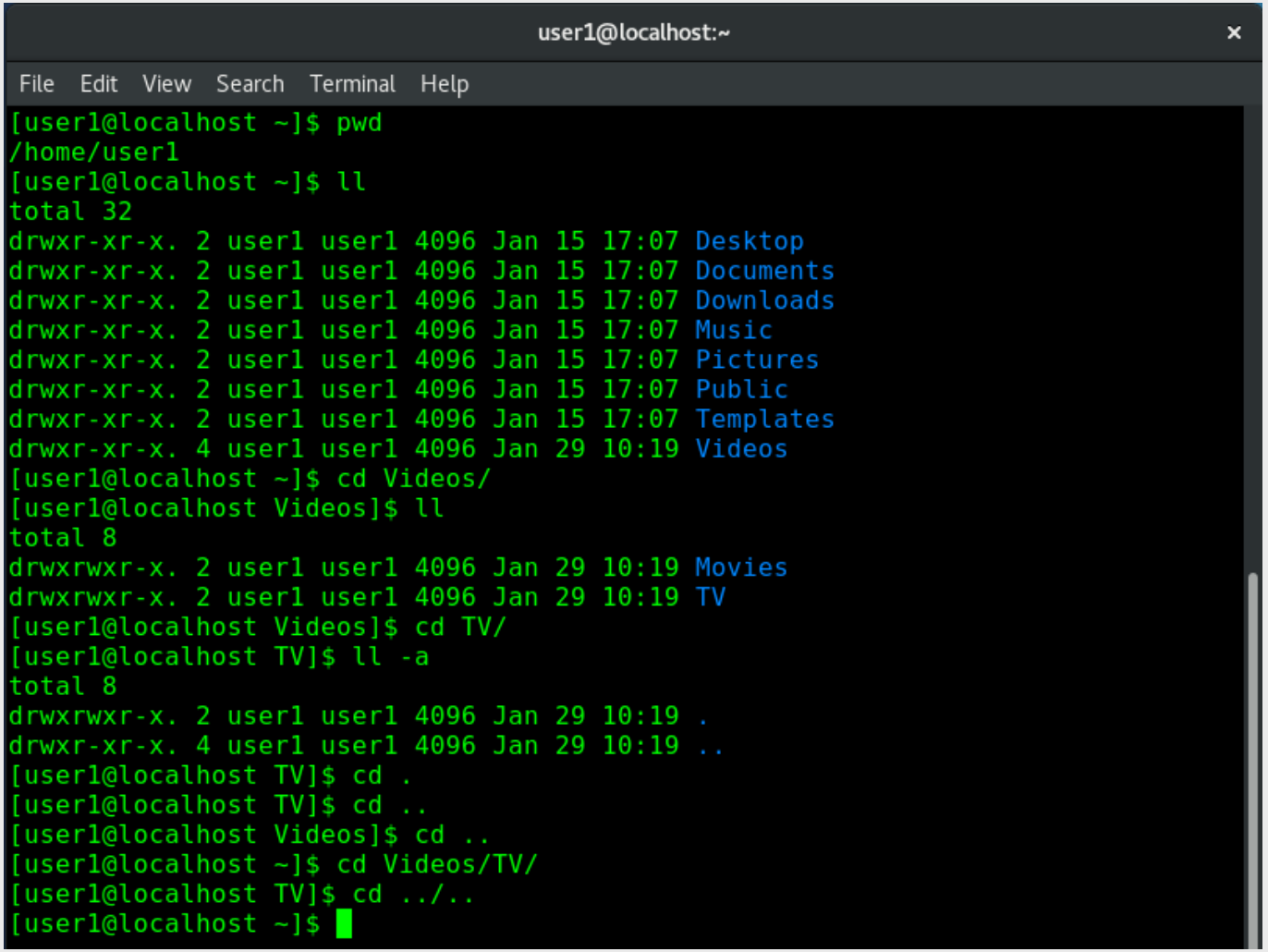
Changing Directories



Changing Directories

- Parent directory
  - Directory one step closer to the root of the tree
  - Referred to by `..` (two dots)
- Current Directory
  - Referred to by `.` (one dot)
- Subdirectory
  - Directory residing within another directory

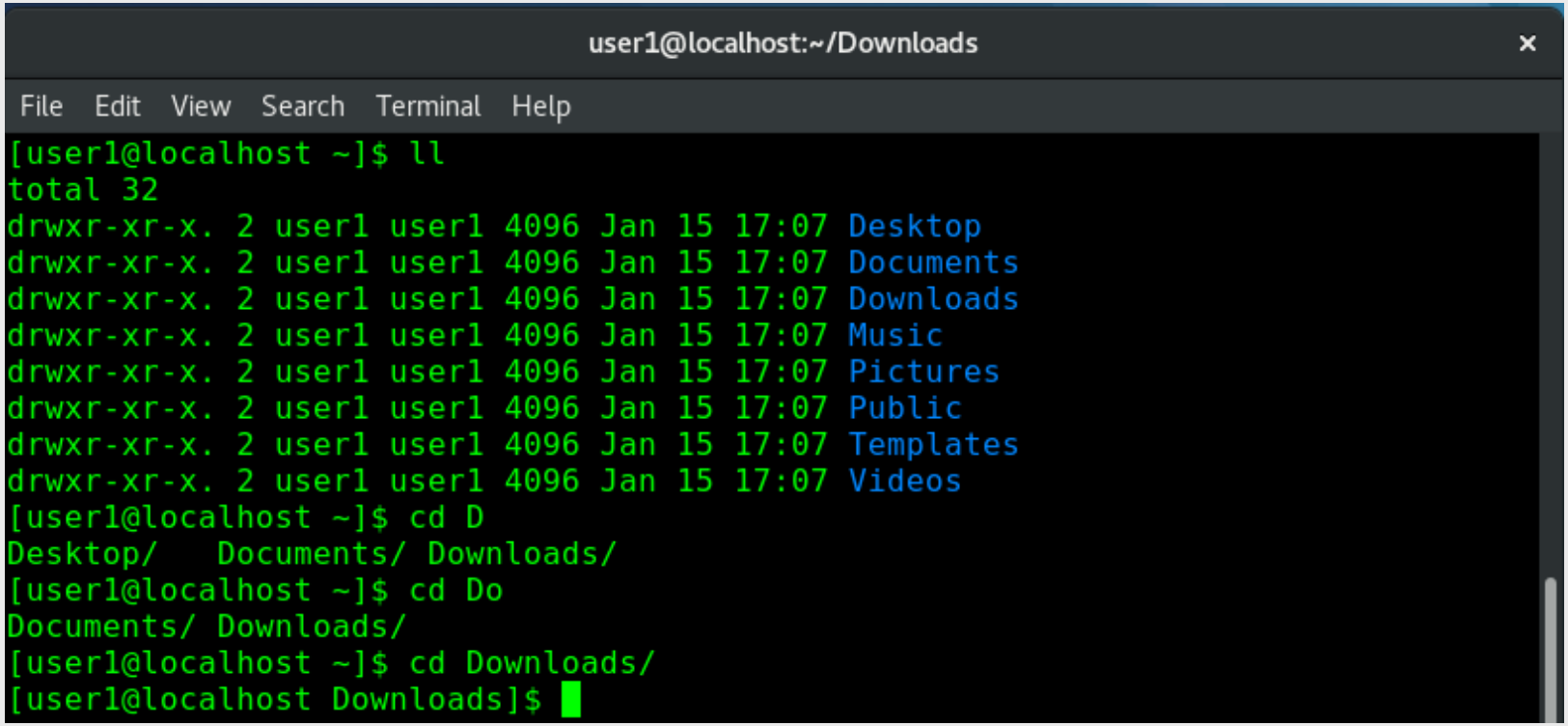
Changing Directories



Changing Directories

- Tab-completion
  - Pressing the Tab key fills in remaining characters of a **unique** filename or directory name
  - BASH shell feature
  - Alerts user if there is more than one possible match
    - Must press tab again

Changing Directories



Viewing Files and Directories

- Point of a directory structure
  - Organize files into an easy-to-use format
- This sections covers various types of files and filenames
  - As well as the different commands used to select filenames for viewing

File Types

- Text files
  - Store information in a readable text format
  - Contain configuration information
- Binary data files
  - Store information associated with executable programs
  - Cannot view the contents

File Types

- Executable program files
  - Also cannot be viewed
- Directory files
  - Serve as placeholders to organize other files
- Linked files
  - Associated with another file
- Special device files
  - Represent system devices

File Types

- Named pipes
  - Identify a channel that passes information between processes
- Socket files
  - Allow a process on another computer to write to a local file

Filenames

- Filename
  - User-friendly identifier given to a file
  - Up to 255 characters (most are short)
  - Can use alphanumeric characters, dash (-), underscore (_), and dot (.)
- Filename extensions
  - Series of identifiers following a dot (.) at the end of a filename
  - Denote the type of file
  - Most files on Linux do not have filename extensions but some do

Filenames

Metacharacter	Description
.c	C programming language source code files
.cc .cpp	C++ programming language source code files
.html .htm	HTML (Hypertext Markup Language) files
.ps	Files formatted for printing with Postscript
.txt	Text files
.tar	Archives files (contains other files within)
.gz .bz2 .Z	Compressed files
.tar.gz .tgz .tar.bz2 .tar.Z	Compressed archive files
.conf .cfg	Configuration files (contain text)

Common [Linux] filename extensions

Filenames

Metacharacter	Description
.so	Shared object (programming library) files
.o	Compiled object files
.pl	Perl (Practical Extraction and Report Language) files
.tcl	Tcl (Tool Command Language) files
.jpg .jpeg .gif .png .tiff .xpm	Binary files that contain graphical images
.sh	Shell scripts (contain text that is executed by the shell)

Common filename extensions

Listing Files

- **`ls`** command
  - Linux command used to list the files in a given directory
  - Most common method for displaying files
  - May pass an argument indicating the directory to be listed
  - Displays all the files in the current directory in columnar format
    - **`-F`** switch: Argument to list file types
    - **`-l`** switch: Argument to list long file listings

Listing Files

- Long listing for each file includes eight components
  - File type character
  - List of permissions (mode of the file)
  - Hard link count
  - Owner
  - Group owner
  - File size
  - Most recent modification time
  - Filename

Listing FIles


```
[root@localhost /]# ls
bin   dev  home  lib64       media  opt   root  sbin  sys  usr
boot  etc  lib   lost+found  mnt    proc  run   srv   tmp  var
[root@localhost /]# ls -F
bin@   dev/  home/  lib64@       media/  opt/   root/  sbin@  sys/  usr/
boot/  etc/  lib@   lost+found/  mnt/    proc/  run/   srv/   tmp/  var/
[root@localhost /]# ls -l
total 62
lrwxrwxrwx.   1 root root     7 Dec 11  2013 bin -> usr/bin
dr-xr-xr-x.   6 root root  1024 Nov 30 12:18 boot
drwxr-xr-x.  20 root root  3300 Jan 23 21:41 dev
drwxr-xr-x. 143 root root 12288 Jan 23 21:41 etc
drwxr-xr-x.   3 root root  4096 Nov 29 18:34 home
lrwxrwxrwx.   1 root root     7 Dec 11  2013 lib -> usr/lib
lrwxrwxrwx.   1 root root     9 Dec 11  2013 lib64 -> usr/lib64
drwx------.   2 root root 16384 Dec 11  2013 lost+found
drwxr-xr-x.   2 root root  4096 Aug  7  2013 media
drwxr-xr-x.   3 root root  4096 Nov 29 18:34 mnt
drwxr-xr-x.   2 root root  4096 Aug  7  2013 opt
dr-xr-xr-x. 194 root root     0 Jan 23 21:41 proc
dr-xr-x---.   9 root root  4096 Jan 23 21:41 root
drwxr-xr-x.  43 root root  1140 Jan 23 21:41 run
lrwxrwxrwx.   1 root root     8 Dec 11  2013 sbin -> usr/sbin
drwxr-xr-x.   2 root root  4096 Aug  7  2013 srv
dr-xr-xr-x.  13 root root     0 Jan 23 21:41 sys
drwxrwxrwt.  13 root root   320 Jan 23 21:41 tmp
drwxr-xr-x.  12 root root  4096 Dec 11  2013 usr
drwxr-xr-x.  21 root root  4096 Jan 23 21:41 var
[root@localhost /]#
```

Listing Files

- Alias
  - Shortcut for a command
  - Can create your own alias for commonly used commands
  - **`alias ll='ls -lrth'`**
  - Must add to .bashrc
    - Hidden file in user's home directory
- **`ll`** command is the alias for **`ls -l`**
  - Gives a long file listing

Listing Files


```
[root@localhost /]# alias
alias cp='cp -i'
alias egrep='egrep --color=auto'
alias fgrep='fgrep --color=auto'
alias grep='grep --color=auto'
alias l.='ls -d .* --color=auto'
alias ll='ls -l --color=auto'
alias ls='ls --color=auto'
alias mv='mv -i'
alias rm='rm -i'
alias which='alias | /usr/bin/which --tty-only --read-alias --show-dot --show-ti
lde'
[root@localhost /]# ll
total 62
lrwxrwxrwx.   1 root root     7 Dec 11  2013 bin -> usr/bin
dr-xr-xr-x.   6 root root  1024 Nov 30 12:18 boot
drwxr-xr-x.  20 root root  3300 Jan 23 21:41 dev
drwxr-xr-x. 143 root root 12288 Jan 23 21:41 etc
drwxr-xr-x.   3 root root  4096 Nov 29 18:34 home
lrwxrwxrwx.   1 root root     7 Dec 11  2013 lib -> usr/lib
lrwxrwxrwx.   1 root root     9 Dec 11  2013 lib64 -> usr/lib64
drwx------.   2 root root 16384 Dec 11  2013 lost+found
drwxr-xr-x.   2 root root  4096 Aug  7  2013 media
drwxr-xr-x.   3 root root  4096 Nov 29 18:34 mnt
drwxr-xr-x.   2 root root  4096 Aug  7  2013 opt
dr-xr-xr-x. 185 root root     0 Jan 23 21:41 proc
dr-xr-x---.   9 root root  4096 Jan 23 21:41 root
drwxr-xr-x.  43 root root  1140 Jan 23 21:41 run
lrwxrwxrwx.   1 root root     8 Dec 11  2013 sbin -> usr/sbin
drwxr-xr-x.   2 root root  4096 Aug  7  2013 srv
dr-xr-xr-x.  13 root root     0 Jan 23 21:41 sys
drwxrwxrwt.  13 root root   320 Jan 23 21:41 tmp
drwxr-xr-x.  12 root root  4096 Dec 11  2013 usr
drwxr-xr-x.  21 root root  4096 Jan 23 21:41 var
```

Types of Files

- **`file`** command
  - Displays the file type of a specified filename
  - Argument indicates what file or files to analyze
  - Identifies between different types of executable files
  - Identifies empty files

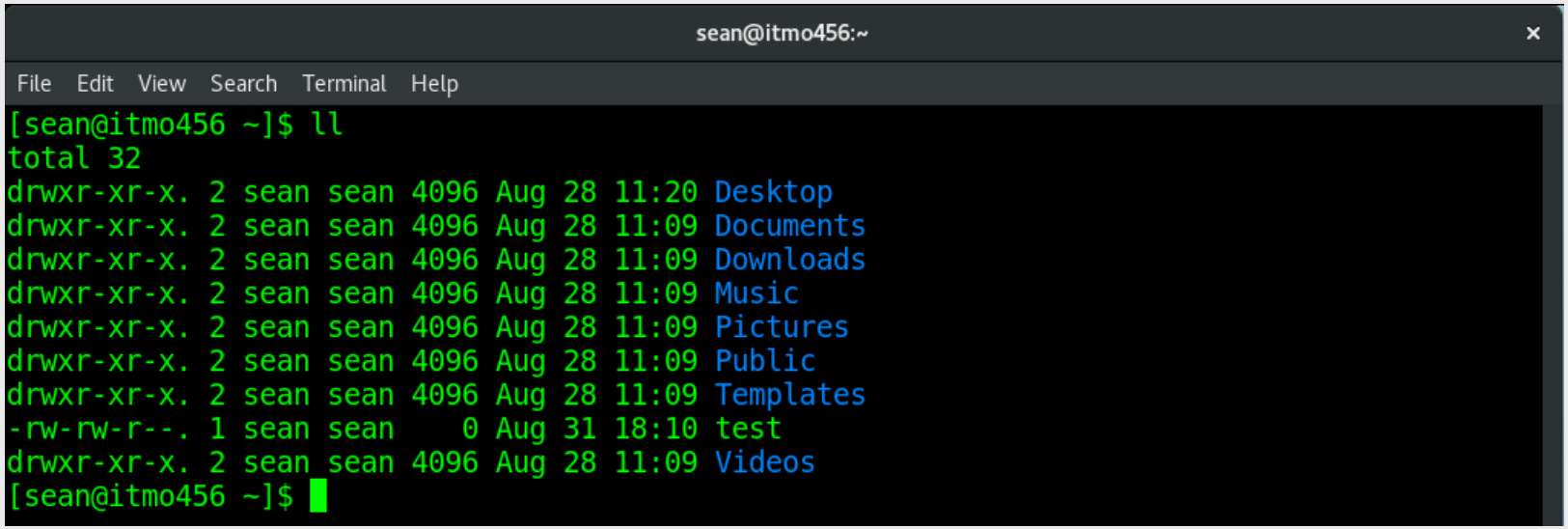
Types of Files


```
[root@localhost ~]# file /bin/bash
/bin/bash: ELF 64-bit LSB executable, x86-64, version 1 (SYSV), dynamically link
ed (uses shared libs), for GNU/Linux 2.6.32, BuildID[sha1]=c9f090657c35c10d6edec
a09f62de9d22060a706, stripped
[root@localhost ~]#
[root@localhost ~]# file /etc/services
/etc/services: ASCII text
[root@localhost ~]#
[root@localhost ~]# file /dev/sr0
/dev/sr0: block special (11/0)
[root@localhost ~]#
[root@localhost ~]# file /bin/unix-lpr.sh
/bin/unix-lpr.sh: POSIX shell script, ASCII text executable
[root@localhost ~]#
[root@localhost ~]# file /dev/tty0
/dev/tty0: character special (4/0)
[root@localhost ~]#
[root@localhost ~]# file /var/mail/sean
/var/mail/sean: empty
[root@localhost ~]#
```

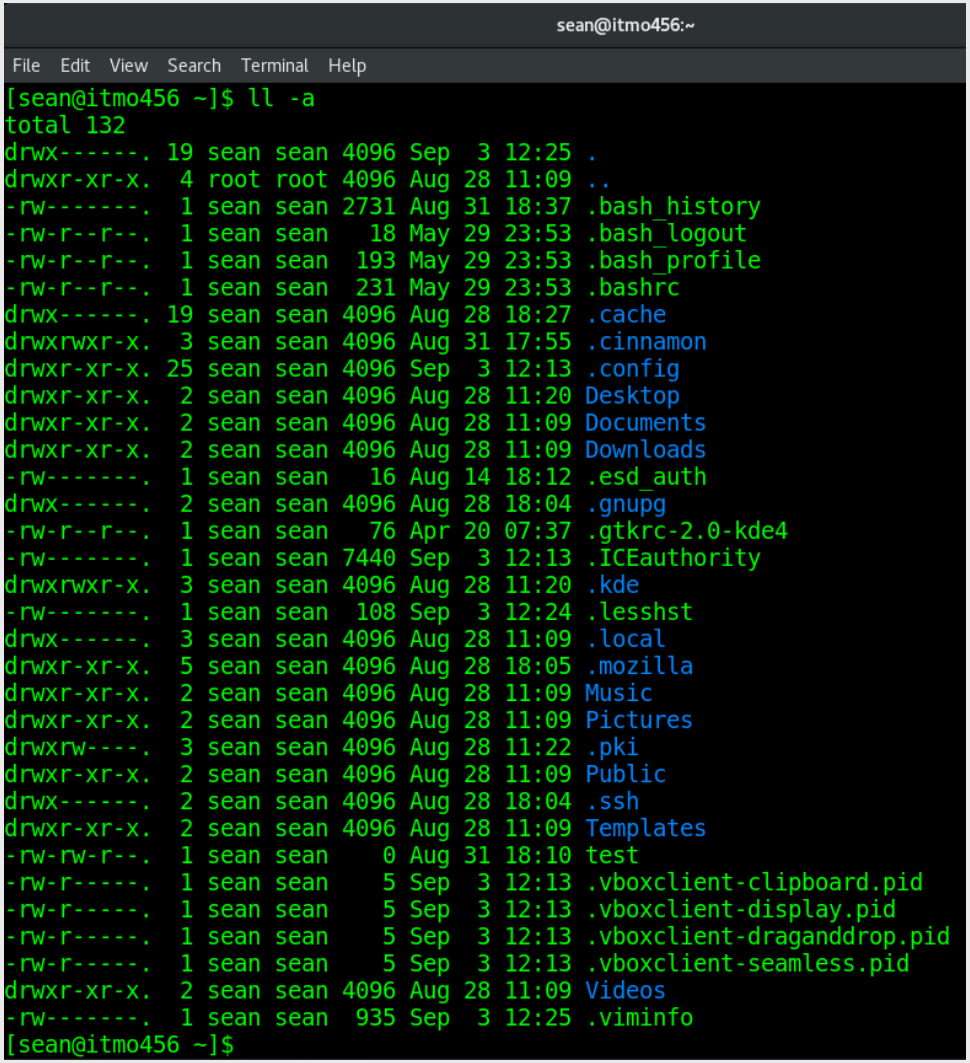
Listing Files

- Hidden files
  - Files not normally displayed to user
  - Configuration files often hidden
  - Filenames start with a dot (`.`)
  - **`ls -a`** command: Displays hidden files
- To view all hidden files and their file types, type:
  - **`ls -a`**
  - **`ll -a`**

Listing Files



Listing Files



Listing Files

Option	Description
-a **--all**	Lists all filenames
-A **--almost-all**	Lists most filenames (excludes the . and .. special files)
-C	Lists filenames in column format
--color=n	Lists filenames without color
-d **--directory**	Lists directory names instead of their contents
-f	Lists all filenames without sorting
-F **--classify**	Lists filenames classified by file type
--full-time	Lists filenames in long format & displays the full modification time
-l	Lists filenames in long format

Common options to the ls command

Listing Files

Option	Description
-lh **-l –human-readable**	Lists filenames in long format with human readable (easy-to-read) file sizes
-r **--reverse**	Lists filenames reverse sorted
-R **--recursive**	Lists filenames in the specified directory and all subdirectories
-s	Lists filenames and their associated filesizes in kilobytes (K)
-S	Lists filenames sorted by file size
-t	Lists filenames sorted by modification date
-U	Lists filenames without sorting
-x	Lists filenames in rows rather than in columns

Common options to the ls command

Wildcard Metacharacters

- Wildcard metacharacter
  - Can simplify commands that specify more than one filename on the command line
  - Interpreted by the shell and can be used with most common Linux filesystem commands
  - Matches certain portions of filenames, or the entire filename itself

Wildcard Metacharacters

Metacharacter	Description
*	Matches 0 or more characters in a filename
?	Matches 1 character in a filename
[aegh]	Matches 1 character in a filename, provided this character is either an a, e, g or h
[a-e]	Matches 1 character in a filename, provided this character is either an a, b, c, d or e
[!a-e]	Matches 1 character in a filename, provided this character is NOT an a, b, c, d or e

Wildcard metacharacters

Wildcard Metacharacters


```
File  Edit  View  Search  Terminal  Help
[user1@localhost Assignments]$ ll labs1*
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs10.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs11.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs12.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs13.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs14.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs15.txt
[user1@localhost Assignments]$ ll labs??.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs01.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs02.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs03.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs04.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs05.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs06.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs07.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs08.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs09.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs10.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs11.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs12.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs13.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs14.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 labs15.txt
[user1@localhost Assignments]$ ll [tb]est??.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 best01.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 best02.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 best03.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 best04.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:11 best05.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test01.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test02.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test03.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test04.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test05.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test06.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test07.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test08.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test09.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test10.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test11.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test12.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test13.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test14.txt
-rw-rw-r--. 1 user1 user1 0 Jan 29 12:14 test15.txt
[user1@localhost Assignments]$
```

Displaying Content of Text Files

- Concatenation
  - Joining of text together to make one larger whole
  - In Linux, words and strings of text are joined together to form a displayed file
- **`cat`** command
  - Linux command used to display (or concatenate) the entire contents of a text file to the screen
  - **`-n`** switch displays line number and contents

Displaying Content of Text Files

- Log files
  - File that contains past system events
- **`tac`** command
  - Linux command that displays a file to the screen beginning with the last line of the file and ending with the first line of the file
  - Makes viewing latest entries of a log file easier

Displaying Content of Text Files


```
[root@localhost ~]# cat /etc/issue
Fedora release 20 (Heisenbug)
Kernel \r on an \m (\l)

[root@localhost ~]# cat -n /etc/issue
     1  Fedora release 20 (Heisenbug)
     2  Kernel \r on an \m (\l)
     3
[root@localhost ~]# tac /etc/issue

Kernel \r on an \m (\l)
Fedora release 20 (Heisenbug)
[root@localhost ~]#
```

Displaying Content of Text Files

- **`head`** command
  - By default, displays the first 10 lines (including blank lines) of a text file to the terminal screen
- **`tail`** command
  - By default, displays the last 10 lines (including blank lines) of a text file to the terminal screen
    - Preferred over **`tac`** command
- For **`head`** and **`tail`** commands
  - Line count includes blank lines
  - Can provide numeric option to specify number of lines to be displayed (e.g., head -2 filename)

Displaying Content of Text Files


```
[root@localhost ~]# head /etc/services
# /etc/services:
# $Id: services,v 1.55 2013/04/14 ovasik Exp $
#
# Network services, Internet style
# IANA services version: last updated 2013-04-10
#
# Note that it is presently the policy of IANA to assign a single well-known
# port number for both TCP and UDP; hence, most entries here have two entries
# even if the protocol doesn't support UDP operations.
# Updated from RFC 1700, ``Assigned Numbers'' (October 1994).  Not all ports
[root@localhost ~]#
[root@localhost ~]# tail /etc/services
3gpp-cbsp       48049/tcp                       # 3GPP Cell Broadcast Service Protocol
isnetserv       48128/tcp                       # Image Systems Network Services
isnetserv       48128/udp                       # Image Systems Network Services
blp5            48129/tcp                       # Bloomberg locator
blp5            48129/udp                       # Bloomberg locator
com-bardac-dw   48556/tcp                       # com-bardac-dw
com-bardac-dw   48556/udp                       # com-bardac-dw
iqobject        48619/tcp                       # iqobject
iqobject        48619/udp                       # iqobject
matahari        49000/tcp                       # Matahari Broker
[root@localhost ~]#
```

Displaying Content of Text Files

- Large text files can not be viewed using the cat command, because the screen will only fit a portion of the file
- **`more`** command
  - Linux command used to display a text file page-by-page and line-by-line
  - Pressing spacebar displays the next page
  - Pressing enter displays the next line
  - Can only scroll down through file
  - Displays a prompt indicating how much of the file is displayed on the screen as %

Displaying Content of Text Files

- **`less`** command
  - Displays a text file page-by-page on the terminal
  - Users may use cursor keys to navigate the file along with space/enter like with **`more`**
  - Better than **`more`**
  - Exit with **`q`**
- The **`more`** and **`less`** commands can also be used in conjunction with output of other commands
  - If output is too large to fit on terminal screen, use "**`|`**" metacharacter, e.g. **`ls -l | less`**

Displaying Binary File Contents

- It is important to employ text file commands as **`cat`**, **`tac`**, **`head`**, **`tail`**, **`more`**, and **`less`** only on files that contain text
  - Otherwise you may find yourself with random output on the terminal screen, or even a dysfunctional screen
- To display a non-text file, typically use the program that created the file

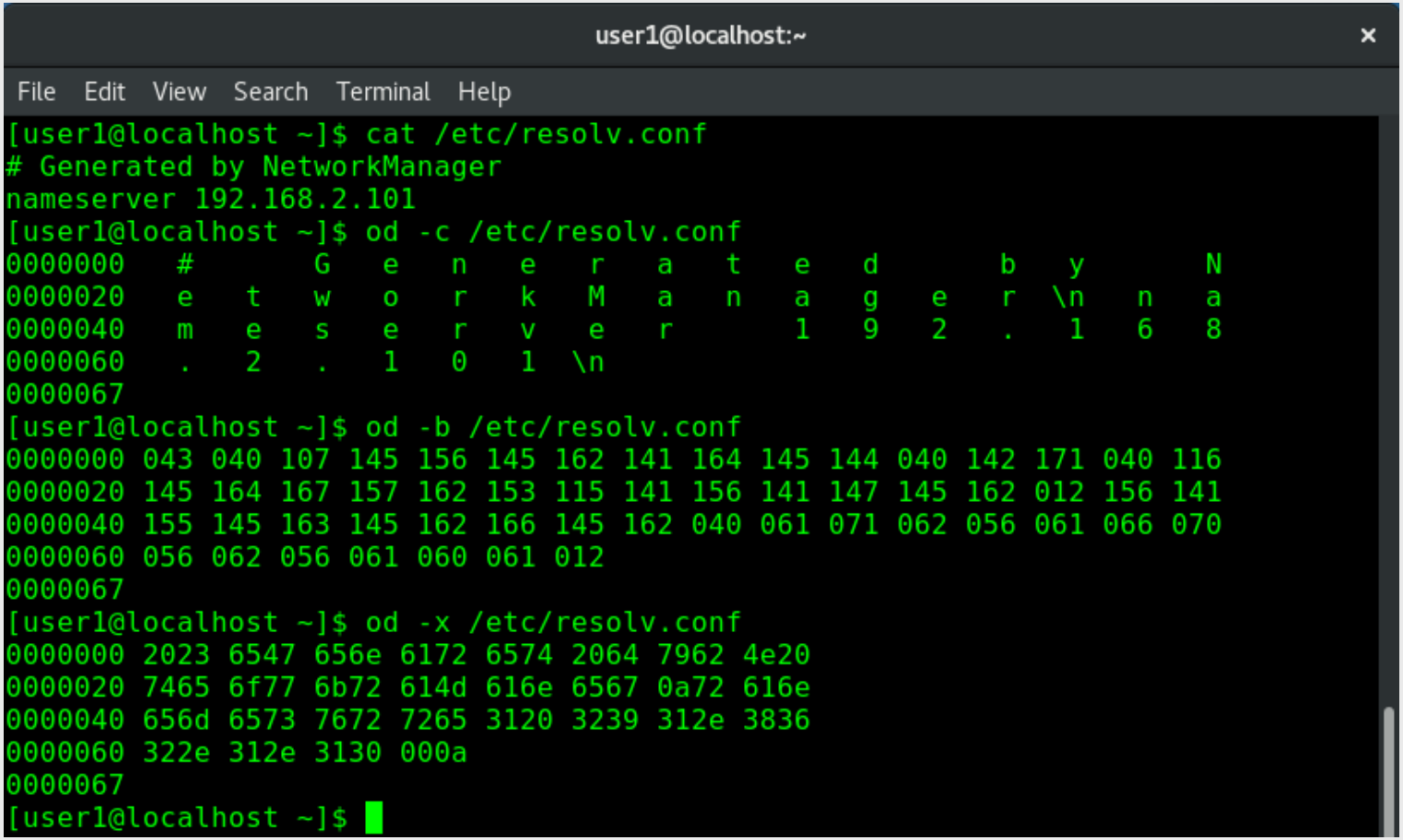
Displaying Binary File Contents

- **`strings`** commands
  - Linux command used to search for and display text characters in a binary file
  - Might indicate purpose of binary file
- **`od`** command
  - Linux command that is used to display the contents of a file in octal format (numeric base 8 format)
  - Safe to use on binary files and text files

Displaying Binary File Contents


```
, +%ld errors
, %d%% packet loss
, time %ldms
%spipe %d
%sipg/ewma %d.%03d/%d.%03d ms
%ld/%ld packets, %d%% loss
ping: recvmsg
ping: bad number of packets to transmit.
ping: bad preload value, should be 1..%d
ping: cannot set preload to value > 3
ping: patterns must be specified as hex digits.
%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x%2x
ping: illegal negative packet size %d.
ping: packet size too large: %d
no answer yet for icmp_seq=%lu
WARNING: probably, rcvbuf is not enough to hold preload.
ping: cannot flood; minimal interval allowed for user is %dms
ping: illegal preload and/or interval
Warning: no SO_TIMESTAMP support, falling back to SIOCGSTAMP
Warning: Failed to set mark %d
Warning: time of day goes back (%ldus), taking countermeasures.
wrong data byte #%d should be 0x%x but was 0x%x
rtt min/avg/max/mdev = %ld.%03ld/%lu.%03ld/%ld.%03ld/%ld.%03ld ms
, min/avg/ewma/max = %ld.%03ld/%lu.%03ld/%d.%03d/%ld.%03ld ms
RMb@A
;*3$"
s20140519
s20140519
[root@localhost ~]# strings /bin/ping
```

Displaying Binary File Contents



Searching for Text within Files

- Text tools
  - commands that search for & manipulate text
- Regular expressions (regex)
  - Text wildcards that ease the search for specific text
  - Match patterns of text within a text document
  - Used by many text tools and programming languages such as **`grep`**

Regular Expressions

- Differences between regular expressions and wildcard metacharacters include:
  - Wildcard metacharacters are interpreted by the shell
    - Regular expressions are interpreted by a text tool program
  - Wildcard metacharacters match characters in filenames (or directory names) on a Linux filesystem
    - Regular expressions match characters within text files on a Linux filesystem

Regular Expressions

- Differences between regular expressions and wildcard metacharacters include (continued):
  - Wildcard metacharacters typically have different definitions than regular expressions
  - There are more regular expressions than wildcard metacharacters
  - Regular expressions are divided into two different categories:
    - Common
    - Extended

Regular Expressions

Regular Expression	Description	Example	Type
*	Matches 0 or more occurrences of the previous character	**lette*** matches lette, letter, letterr, letterrrr, letterrrrrr etc.	Common
?	Matches 0 or 1 occurrences of the previous character	**lette?** matches lette, letter	Extended
+	Matches 1 or more occurrences of the previous character	**letter+** matches letter, letterr, letterrrr, letterrrrrr etc.	Extended
. (period)	Matches 1 character of any type	**letter.** matches lettera, letterb, letterc, letter1, letter2, letter3, etc.	Common
[…]	Matches 1 character from the range specified within the braces	**Letter[1238]** matches letter1, letter2, letter3, & letter8	Common

Regular expressions

Regular Expressions

Regular Expression	Description	Example	Type
[!…]	Matches 1 character NOT from the range specified within the braces	**letter[!1238]** matches letter4, letter5, letter6, lettera, letterb, etc. (any character except 1,2,3 or 8)	Common
{}	Matches a specific range or number of the previous character	**letter{3}** matches letterrr **letter{2,4}** matches letterr, letterrr and letterrrr	Extended
^	Matches the following characters if they are the first characters on the line	**^letter** matches letter if letter is the first set of characters in the line	Common
$	Matches previous characters if they are the last characters on the line	**letter$** matches letter if letter is the last set of characters in the line	Common
(…|…)	Matches either of the two sets of characters	**(mother|father)** matches the word mother or father	Extended

Table 3-4: Regular expressions

Searching for Text within Files

- Text tools that use regular expressions include:
  - **grep**
  - **awk**
  - **sed**
  - **vi**
  - **emacs**

The grep Command

- **`grep`**
  - Stands for *Global Regular Expression Print*
  - Used to display lines in a text file that match a certain common regular expression
- **`egrep`** command (**`grep -E`**)
  - Display lines of text that match extended regular expressions
- **`fgrep`** command (**`grep -F`**)
  - Does not interpret any regular expressions and consequently returns results much faster than the egrep command

The grep Command

- **`grep`** requires two arguments
  - Text to search for
    - Can use regular expressions
  - Files in which to search
- **`grep`** is case sensitive
  - For case-insensitive search, use **`-i`** option
- **`grep`** matches patterns of text, ignoring division into words
  - To search only for occurrences of a word, surround it by space characters

The grep Command


```
[root@localhost ~]# grep smtp /etc/services
smtp            25/tcp          mail
smtp            25/udp          mail
urd             465/tcp         smtps   # URL Rendesvous Directory for SSM / SMTP over SSL (TLS)
rsmtp           2390/tcp                # RSMTP
rsmtp           2390/udp                # RSMTP
[root@localhost ~]#
[root@localhost ~]# egrep "(^http|^https)" /etc/services
http            80/tcp          www www-http    # WorldWideWeb HTTP
http            80/udp          www www-http    # HyperText Transfer Protocol
http            80/sctp                         # HyperText Transfer Protocol
https           443/tcp                         # http protocol over TLS/SSL
https           443/udp                         # http protocol over TLS/SSL
https           443/sctp                        # http protocol over TLS/SSL
http-mgmt       280/tcp                 # http-mgmt
http-mgmt       280/udp                 # http-mgmt
http-rpc-epmap  593/tcp                 # HTTP RPC Ep Map
http-rpc-epmap  593/udp                 # HTTP RPC Ep Map
httpx           4180/tcp                # HTTPX
httpx           4180/udp                # HTTPX
http-wmap       8990/tcp                # webmail HTTP service
http-wmap       8990/udp                # webmail HTTP service
https-wmap      8991/tcp                # webmail HTTPS service
https-wmap      8991/udp                # webmail HTTPS service
[root@localhost ~]#
```

The grep Command

sean@itmo456:~

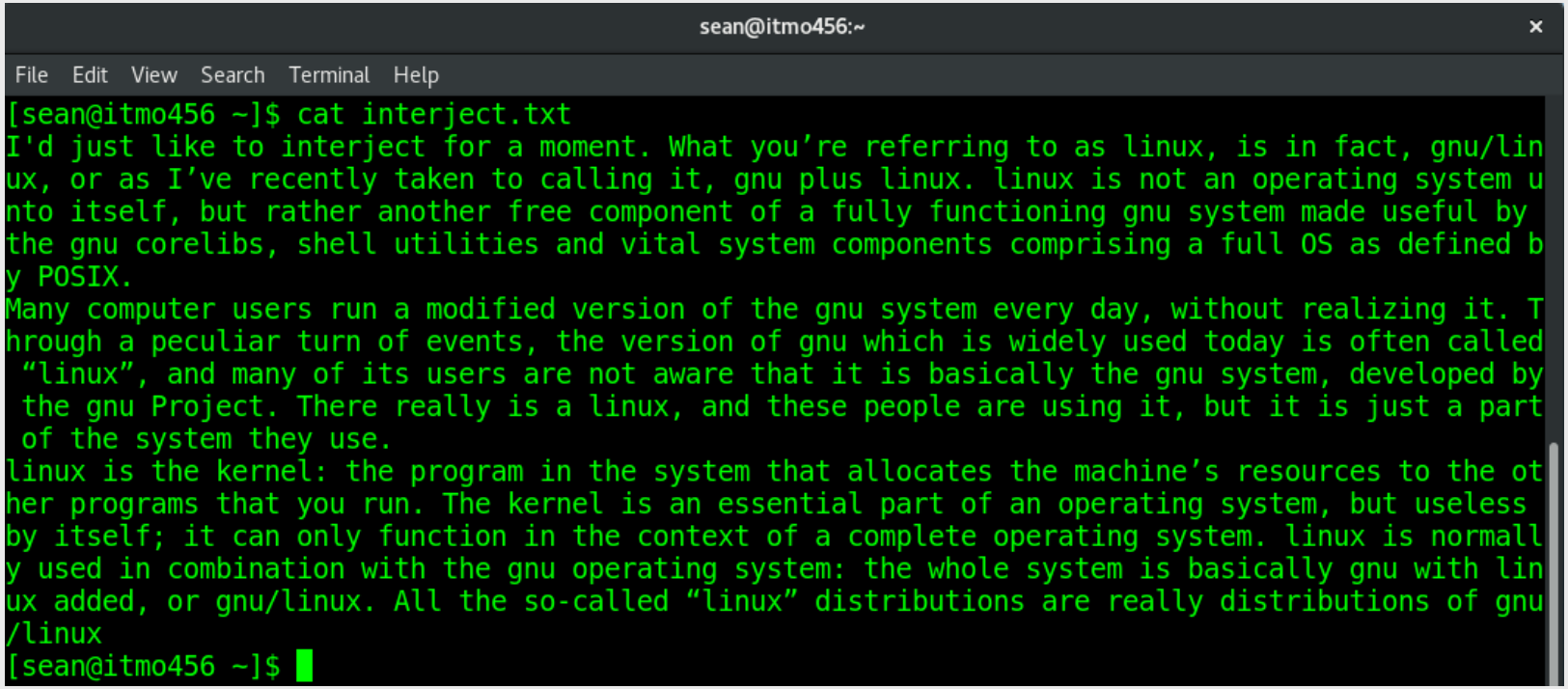
File Edit View Search Terminal Help

```
[sean@itmo456 ~]$ egrep -i "(http|https)" /etc/services
#       http://www.iana.org/assignments/port-numbers
http            80/tcp          www www-http    # WorldWideWeb HTTP
http            80/udp          www www-http    # HyperText Transfer Protocol
http            80/sctp                         # HyperText Transfer Protocol
https           443/tcp                         # http protocol over TLS/SSL
https           443/udp                         # http protocol over TLS/SSL
https           443/sctp                        # http protocol over TLS/SSL
gss-http        488/tcp
gss-http        488/udp
webcache        8080/tcp        http-alt        # WWW caching service
webcache        8080/udp        http-alt        # WWW caching service
http-mgmt       280/tcp                 # http-mgmt
http-mgmt       280/udp                 # http-mgmt
http-rpc-epmap  593/tcp                 # HTTP RPC Ep Map
http-rpc-epmap  593/udp                 # HTTP RPC Ep Map
oob-ws-http     623/tcp                 # DMTF out-of-band web services management protocol
oob-ws-https    664/tcp                 # DMTF out-of-band secure web services management proto
col
multiling-http  777/tcp                 # Multiling HTTP
multiling-http  777/udp                 # Multiling HTTP
```

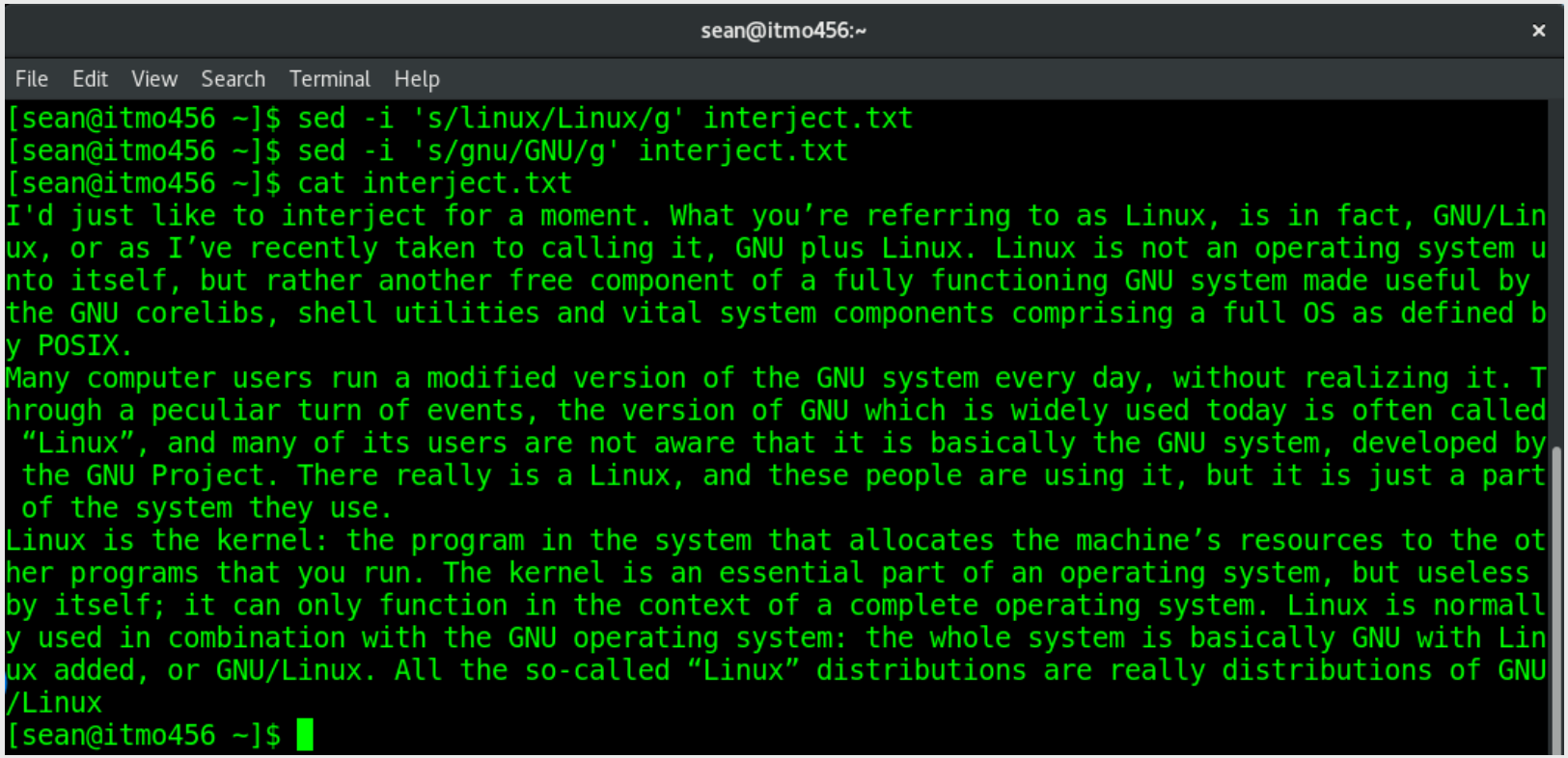
The sed Command

- Stream Editor
- Supports regular expressions
- Used to perform basic text transformations on an input stream
- Basic substitution usage
  - `sed 's/regexp/replacement/g' inputFileName > outputFileName`
  - `s stands for substitute`
  - `g stands for global (more than one occurance)`

The sed Command



The sed Command



The awk Command

- Programming language designed for text processing and typically used as a data extraction and reporting tool
- Developed in 1977 by Alfred Aho, Peter Weinberger, and Brian Kernighan
  - from whose initials the language takes its name
- GNU AWK revised and expanded in 85-88 by Richard Stallman and team

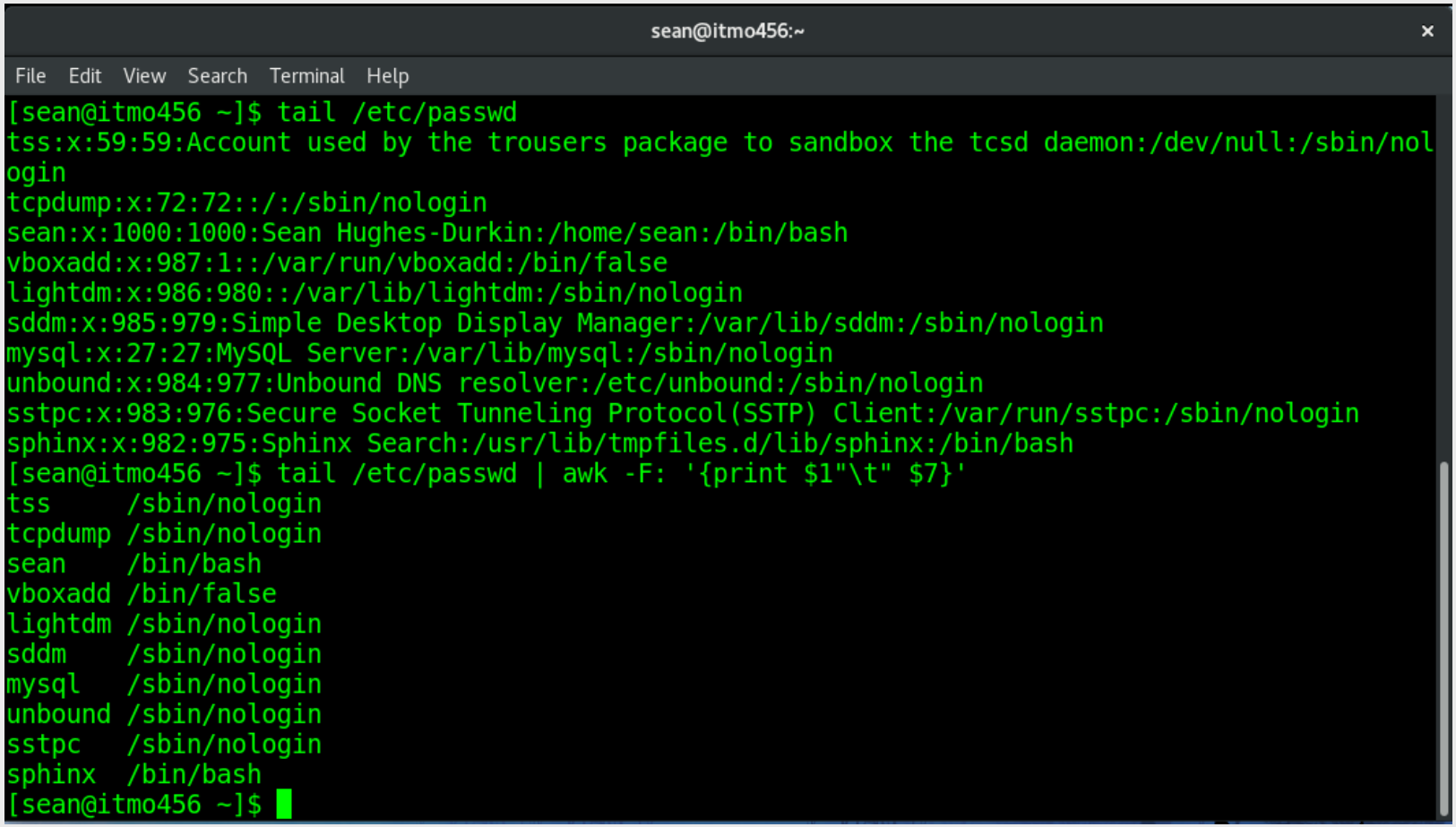
The awk Command

- **AWK** is a language for processing text files
- A file is treated as a sequence of records, and by default each line is a record
- Each line is broken up into a sequence of fields, so we can think of the first word in a line as the first field, the second word as the second field, and so on

The awk Command

- Built-in variables
  - Awk's built-in variables include the field variables: $1, $2, $3, and so on ($0 represents the entire record)
  - They hold the text or values in the individual text-fields in a record
  - Other variables exist

The awk Command



Editing Text Files

- Most system configuration is stored in text files
  - As is commonly accessed information such as e-mail and program source code
- Most Linux distributions come with several text editors
- Text editors come in two varieties:
  - Editors that can be used on the command line
  - Editors that can be used in a GUI

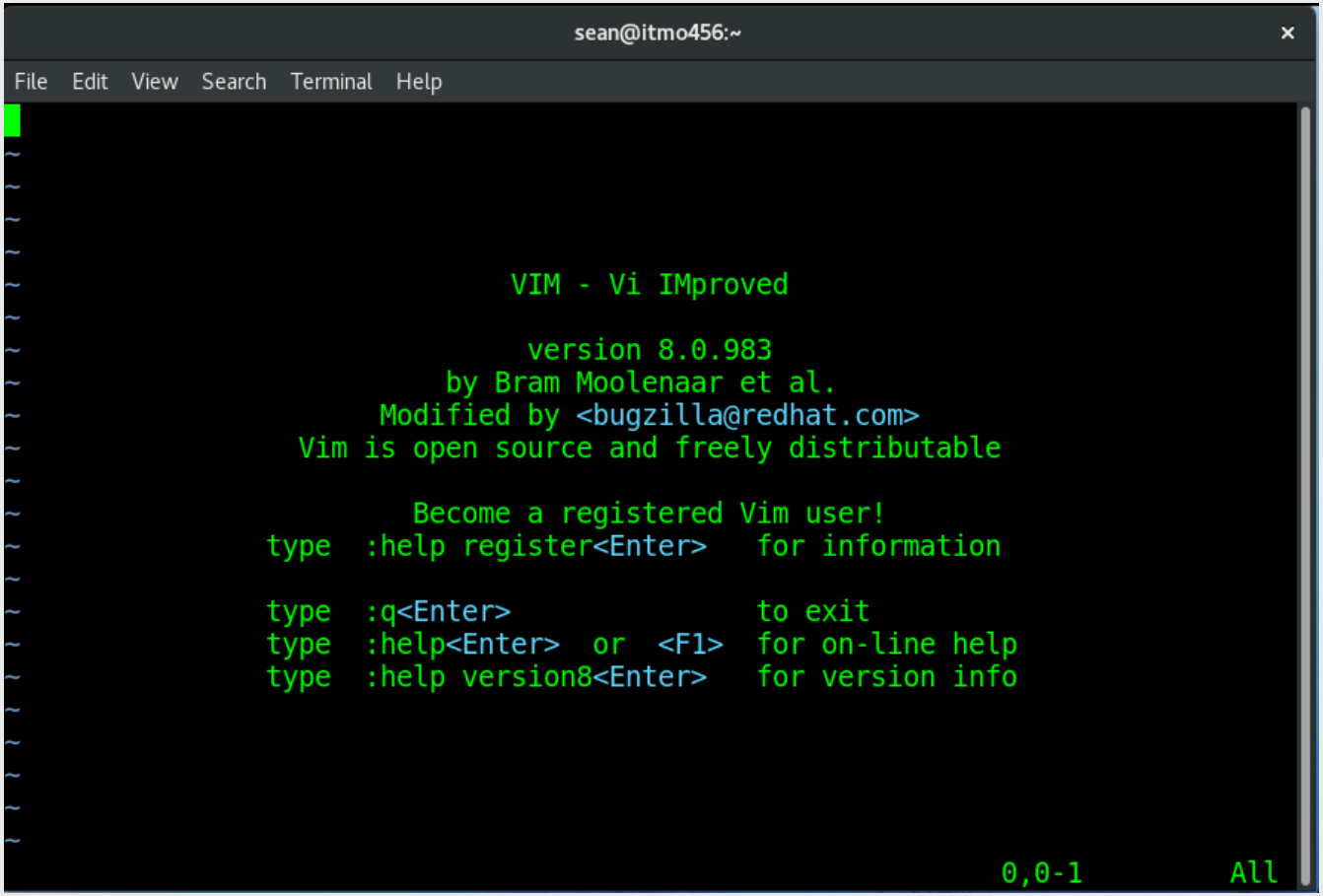
The vi Editor

- One of the oldest and most popular visual text editors available for UNIX operating systems
  - Its Linux equivalent (known as **`vim`**—"vi improved") is standard on almost every Linux distribution as a result
- Advantage is portability
  - Used on Unix and Linux
  - Can perform over 1000 different functions for the user

The vi Editor

- The vi editor is a bi-modal editor as it functions in one of two modes:
  - Command mode
    - Perform text editing tasks not related to inserting text into the document
    - Delete text, copy text, save changes, exit **`vi`**
  - Insert mode
    - Insert text into the document but nothing else
    - Press Esc key to return to command mode
- User environment is customizable

The vi Editor



The vi Editor

Key	Description
i	Changes to insert mode and places the cursor before the current character for entering text
a	Changes to insert mode and places the cursor after the current character for entering text
o	Changes to insert mode and opens up a new line underneath the current line for entering text
Shift-I	Changes to insert mode and places the cursor at the beginning of the current line for entering text
Shift-A	Changes to insert mode and places the cursor at the end of the current line for entering text
Shift-O	Changes to insert mode and opens up a new line above the current line for entering text
[Esc]	Changes back to command mode while in insert mode

vi common keys used to change to/from insert mode

The vi Editor

Key	Description
w, W, e, E	Moves the cursor forward one word
b, B	Moves the cursor backward one word
53G	Moves the cursor to line 53
G	Moves the cursor to the last line in the document
0,^	Moves the cursor to the beginning of the line
$	Moves the cursor to the end of the line
x	Deletes the character the cursor is on
3x	Deletes three characters starting from the character the cursor is on
dw	Deletes one word starting from the character the cursor is on
d3w, 3dw	Deletes three words starting from the character the cursor is on
dd	Deletes the whole line starting from the line the cursor is on
d3d, 3dd	Deletes three whole lines starting from the line the cursor is on
d$	Deletes from the cursor character to the end of the current line
d^, d0	Deletes from the cursor character to the beginning of the current line

Key combinations commonly used in vi command mode

The vi Editor

Key	Description
gg	Moves the cursor to the beginning of the document
yw	Copies one word (starting from the character the cursor is on) into a temporary buffer in memory for later use
y3w, 3yw	Copies three words (starting from the character the cursor is on) into a temporary buffer in memory for later use
yy	Copies the current line into a temporary buffer in memory for later use
y3y, 3yy	Copies three lines (starting from the current line) into a temporary buffer in memory for later use
y$	Copies the current line from the cursor to the end of the line into a temporary buffer in memory for later use
y^, y0	Copies the current line from the cursor to the beginning of the line into a temporary buffer in memory for later use
p	Pastes the contents of the temporary memory buffer underneath the current line

Key combinations commonly used in vi command mode

The vi Editor

Key	Description
P	Pastes the contents of the temporary memory buffer above the current line
J	Joins the line underneath the current line to the current line
[Ctrl]-g	Displays current line statistics
u	Undoes the last function (undo)
.	Repeats the last function (repeat)
/pattern	Searches for the first occurrence of the pattern in the forward direction
?pattern	Searches for the first occurrence of the pattern in the reverse direction
n	Repeats the previous search in the forward direction
N	Repeats the previous search in the reverse direction

Key combinations commonly used in vi command mode

The vi Editor

Function	Description
:q	Quits from the vi editor if no changes were made
:q!	Quits from the vi editor and does not save any changes
:wq	Save any changes to the file and quits from the vi editor
:w filename	Saves the current document to a file called filename
:!date	Executes the date command using a BASH shell
:r !date	Reads the output of the date command into the document under the current line
:r filename	Reads the contents of the text file called filename into the document under the current line
:set all	Displays all vi environment settings
:set ……	Sets a vi environment setting to a certain value
:s/the/THE/g	Searches for the regular expression "the" and replaces each occurrence globally throughout the current line with the word "THE"
:1,$ s/the/THE/g	Searches for the regular expression "the" and replaces each occurrence globally from line 1 to the end of the document with the word "THE"

Key combinations commonly used at the vi command mode : prompt

Other Common Text Editors

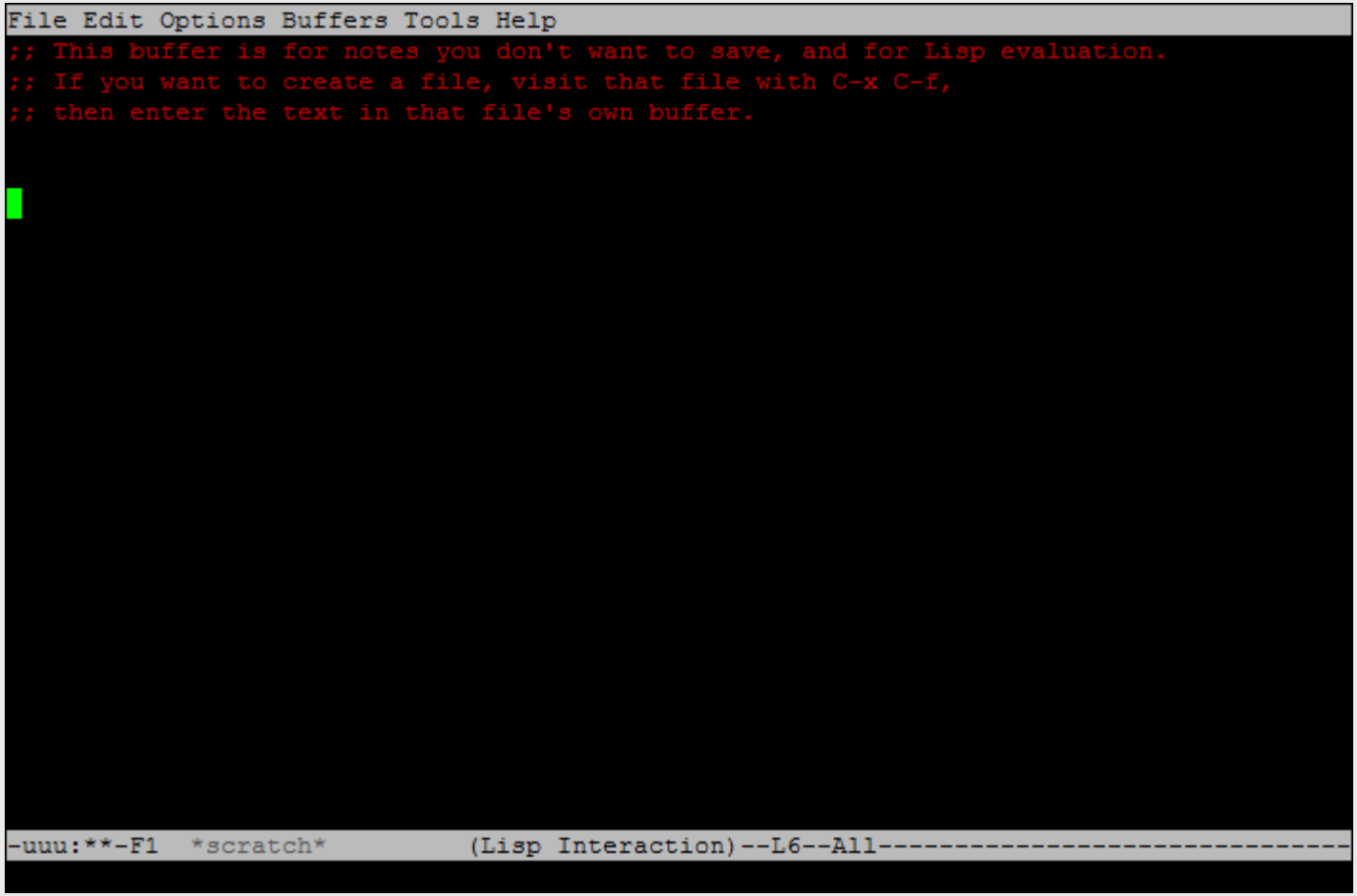
- **`emacs`** (Editor MAcroS) editor
  - **`vi`** alternative with comparable functionality
  - Like **`pico`**, uses the **Ctrl** combined with certain letters to perform special functions
  - Supports hundreds of keyboard functions like **`vi`**
  - To install, type: **`yum install emacs`**

Other Common Text Editors

Key	Description
[Ctrl]-a	Moves the cursor to the beginning of the line
[Ctrl]-e	Moves the cursor to the end of the line
[Ctrl]-h	Displays emacs documentation
[Ctrl]-d	Deletes the current character
[Ctrl]-k	Deletes from the cursor position to the end of the line
[Esc]-d	Deletes the current word
[Ctrl]-x + [Ctrl]-c	Exits the emacs editor
[Ctrl]-x + [Ctrl]-s	Saves the current document
[Ctrl]-x + [Ctrl]- w	Saves the current document as a new filename
[Ctrl]-x +u	Undoes the last change

Keyboard functions commonly used in the GNU emacs editor

Other Common Text Editors

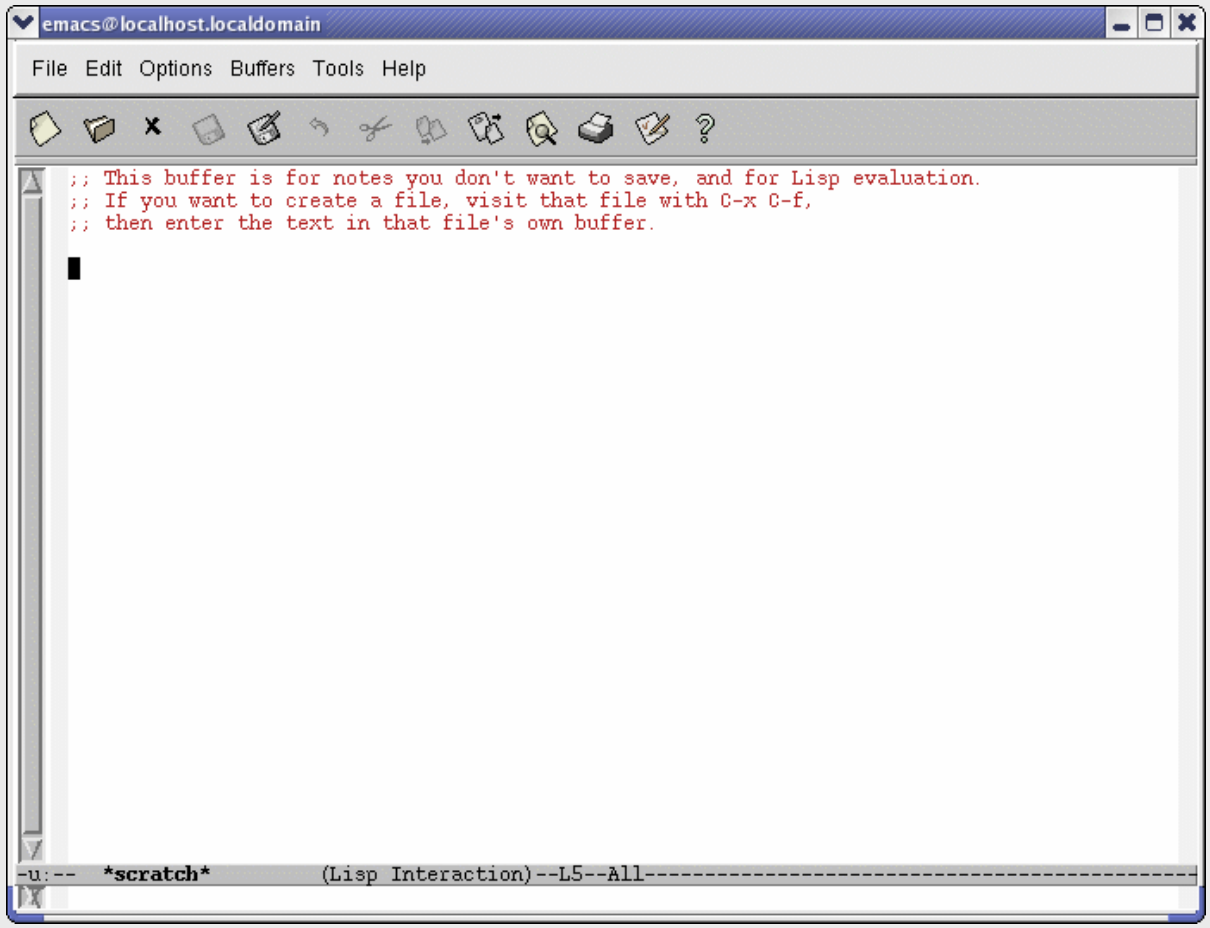


The GNU emacs text editor

Other Common Text Editors

- **`emacs`** not easy to use
  - You must memorize several key combinations to work effectively
- Xemacs editor
  - Version of emacs that runs in the KDE or GNOME GUI environments

Other Common Text Editors



The GNU emacs (Xemacs) text editor

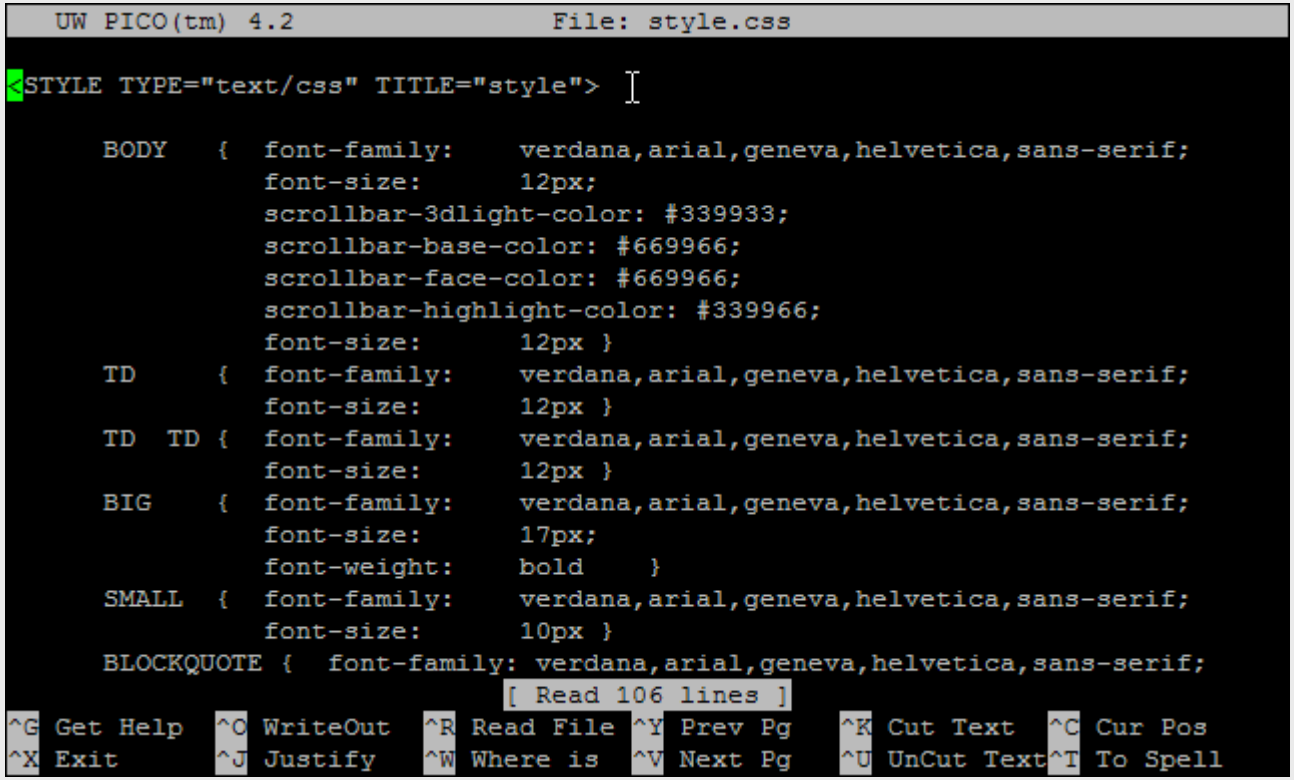
Other Common Text Editors

- **`pico`** (PIne COmposer) editor
  - By far, easiest alternative to vi
  - Part of the Pine email suite from The University of Washington
  - Now rarely included with Linux distros
- **`nano`** (GNU nano editor)
  - Designed to be a free replacement for Pico
  - Aims to "emulate Pico as closely as possible and perhaps include extra functionality"

Other Common Text Editors

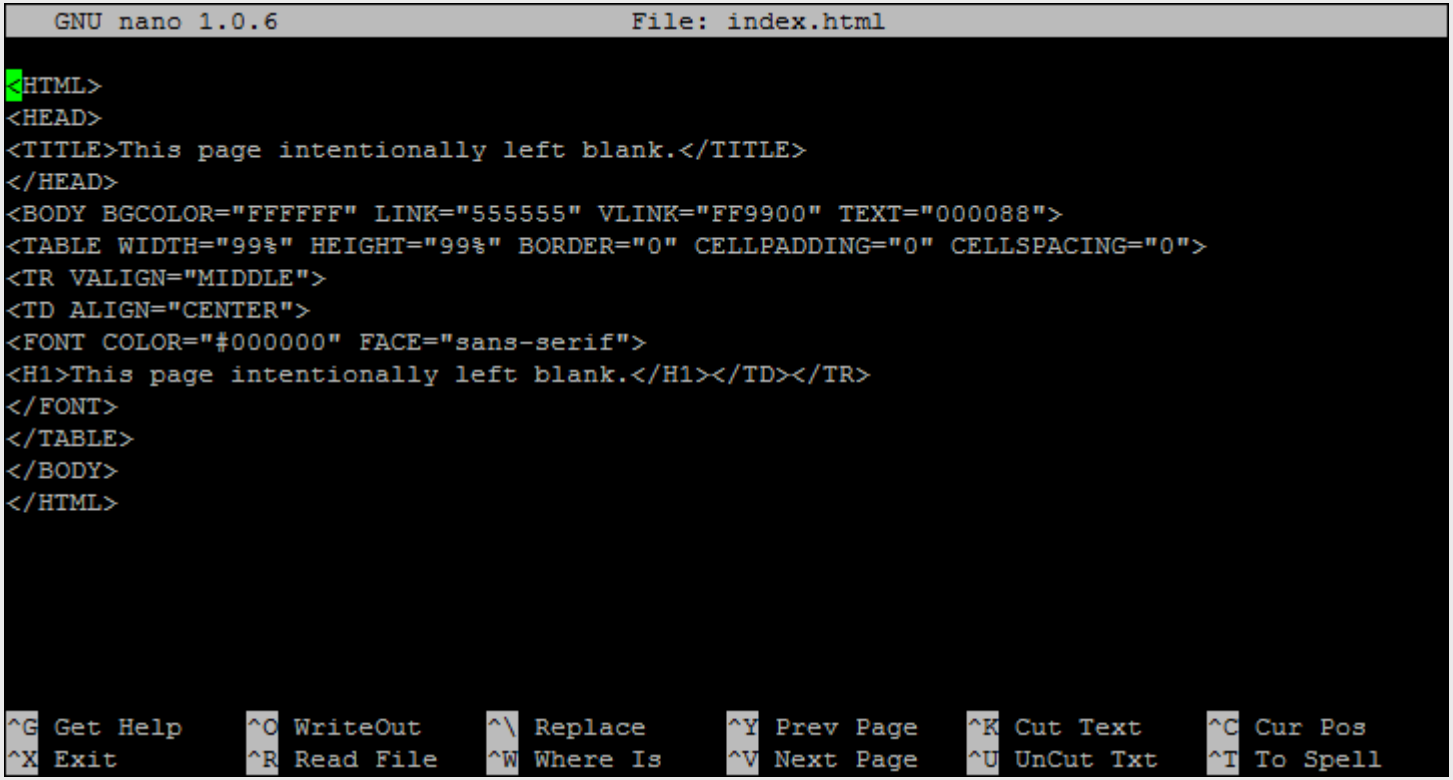
- **nano**
  - uses Ctrl key combinations for performing functions
- Very basic and easy-to-use
  - **Ctrl** key combinations listed at the bottom of the screen
- To install, type: **`yum install nano`**

Other Common Text Editors



pico Editor

Other Common Text Editors

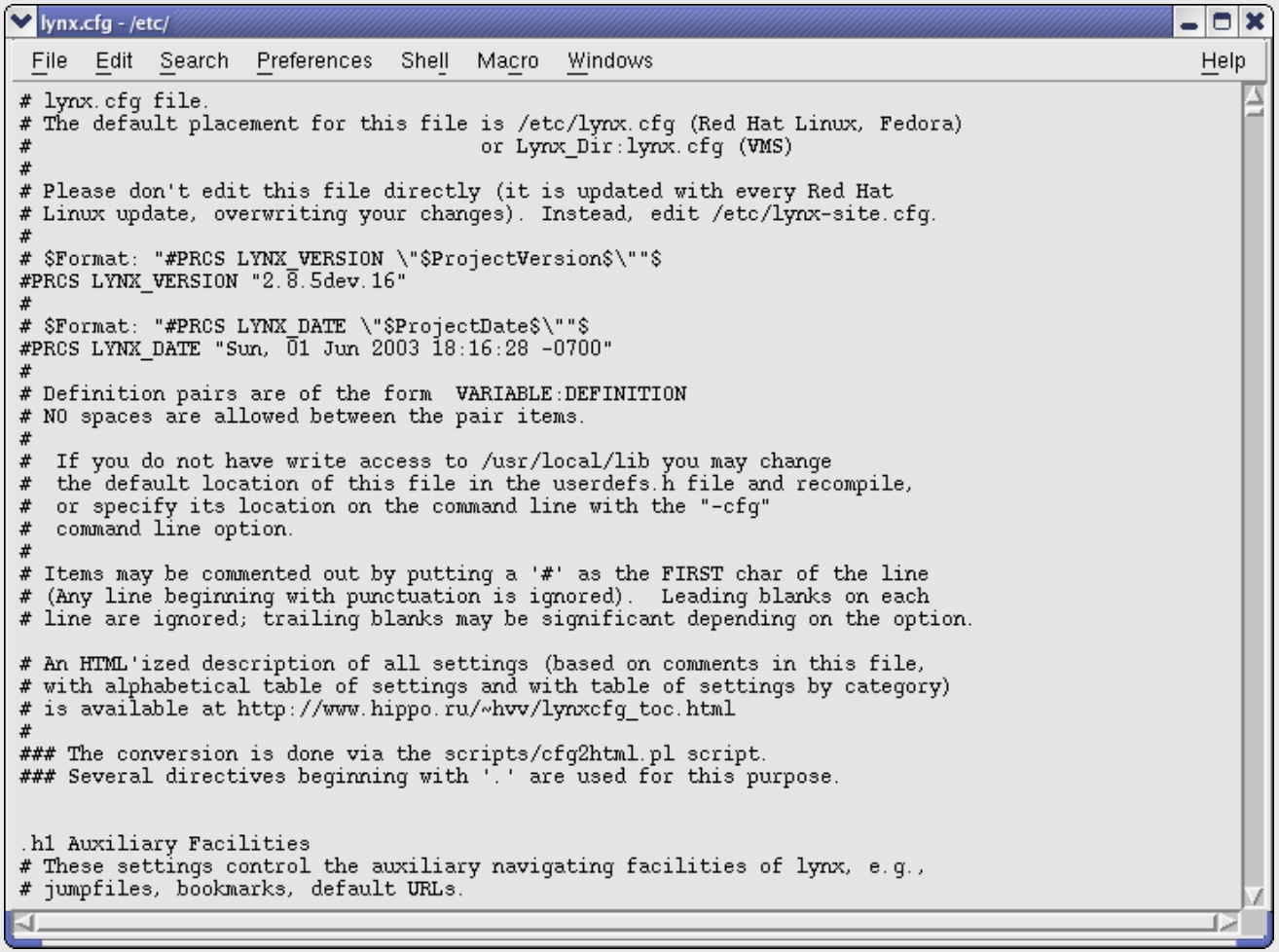


nano Editor

Other Common Text Editors

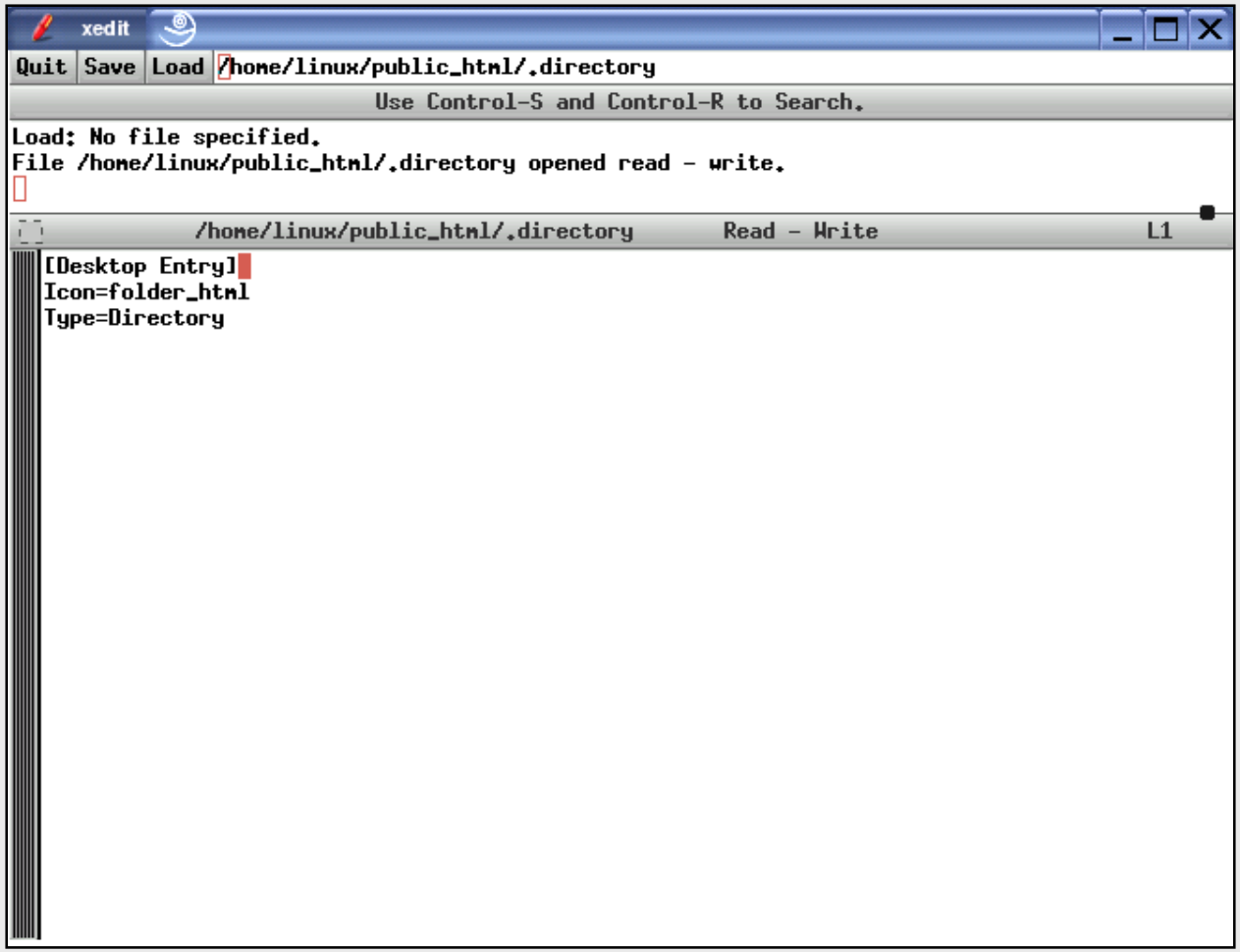
- The xemacs editor may not be available in every Linux distribution that contains a GUI environment
- Commonly used graphical text editors available in most Linux distributions:
  - Nedit editor
  - Xedit editor

Other Common Text Editors



The nedit text editor

Other Common Text Editors



The xedit text editor

Other Common Text Editors

- Gedit editor
  - Text editor for the GNOME desktop
- Kedit editor
  - Text editor for the KDE desktop
  - Often now replaced by Kate or KWrite
- The gedit and kedit editors are similar to nedit yet offer more functionality

Other Common Text Editors

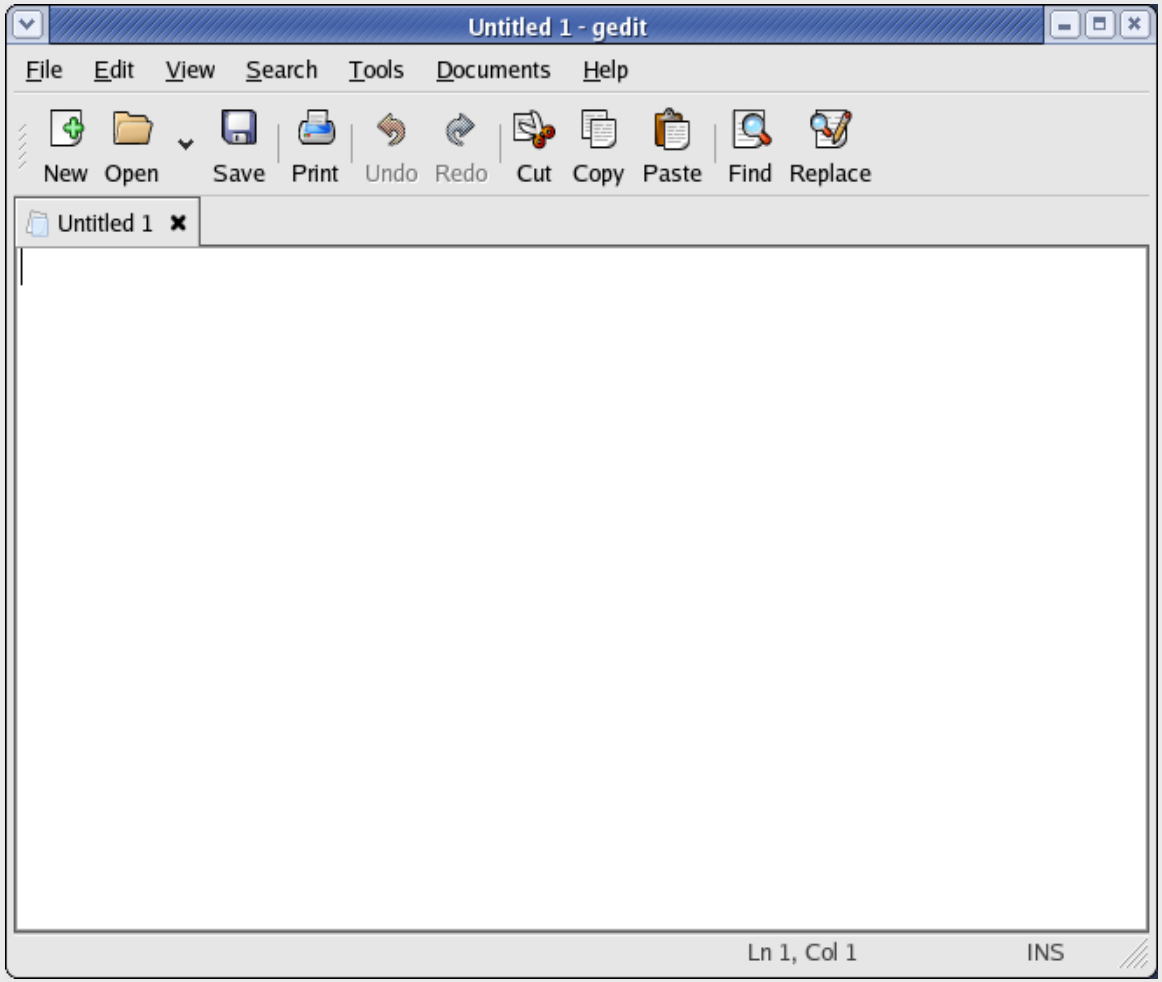
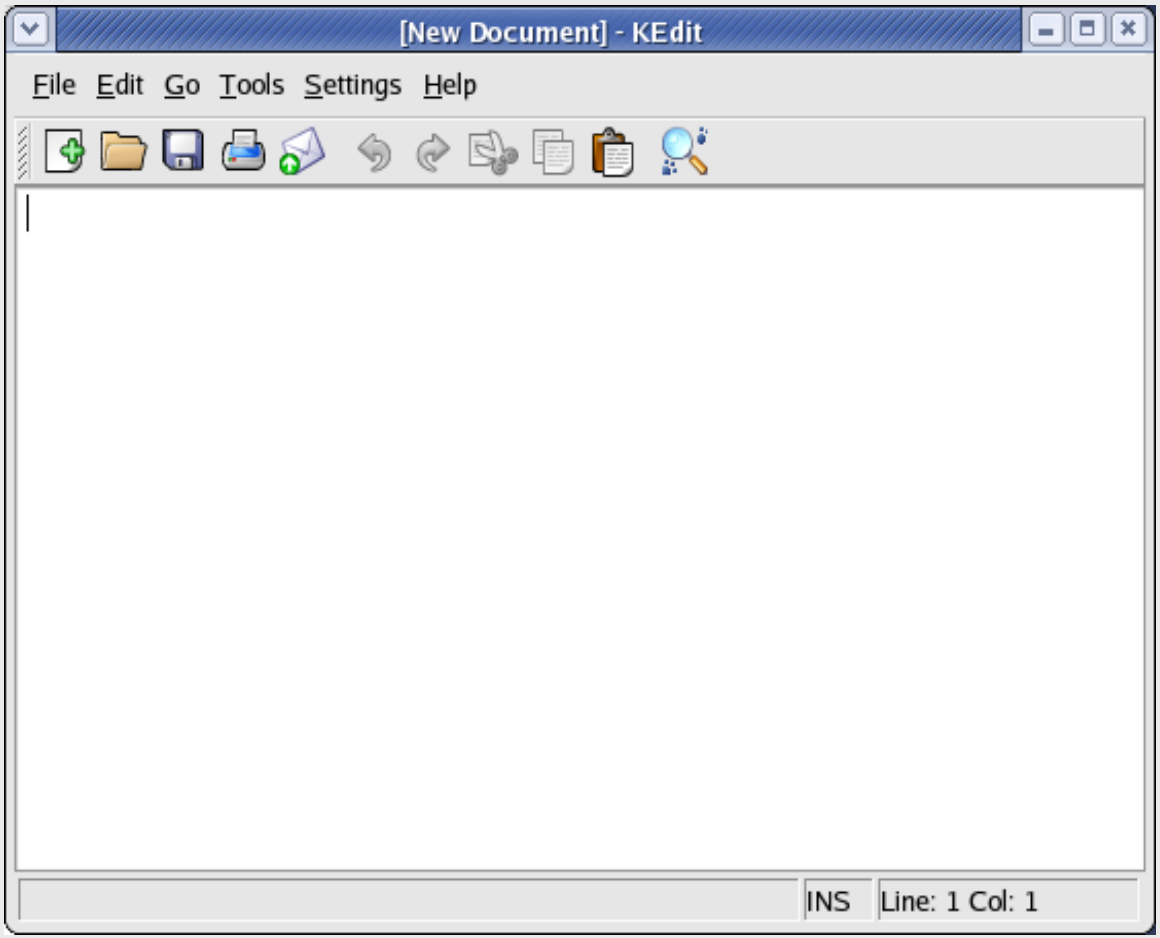


Figure 3-4: The gedit text editor

Other Common Text Editors



The kedit text editor

Summary

- The Linux filesystem is arranged hierarchically using a series of directories to store files, and the location of these directories and files can be described using absolute or relative pathnames
- There are many types of files that may exist on the Linux filesystem
  - text files, binary data, executable programs, directories, linked files, special device files

Summary

- The ls command can be used to view filenames and offers a wide range of options to modify this view
- Wildcard metacharacters can be used to simplify the selection of several files when using common Linux file commands
- Text files are the common file type whose contents may be viewed by several utilities such as **`head`**, **`tail`**, **`cat`**, **`tac`**, **`more`**, and **`less`**

Summary

- Regular expressions can be used to specify certain patterns of text when used with certain programming languages and text tools such as **`grep`**
- **`vi`** (vim) is a powerful, bi-modal text editor that is standard on most UNIX and Linux systems

The End...

- Questions?