CS631 - Advanced Programming in the UNIX Environment
–
UNIX development tools

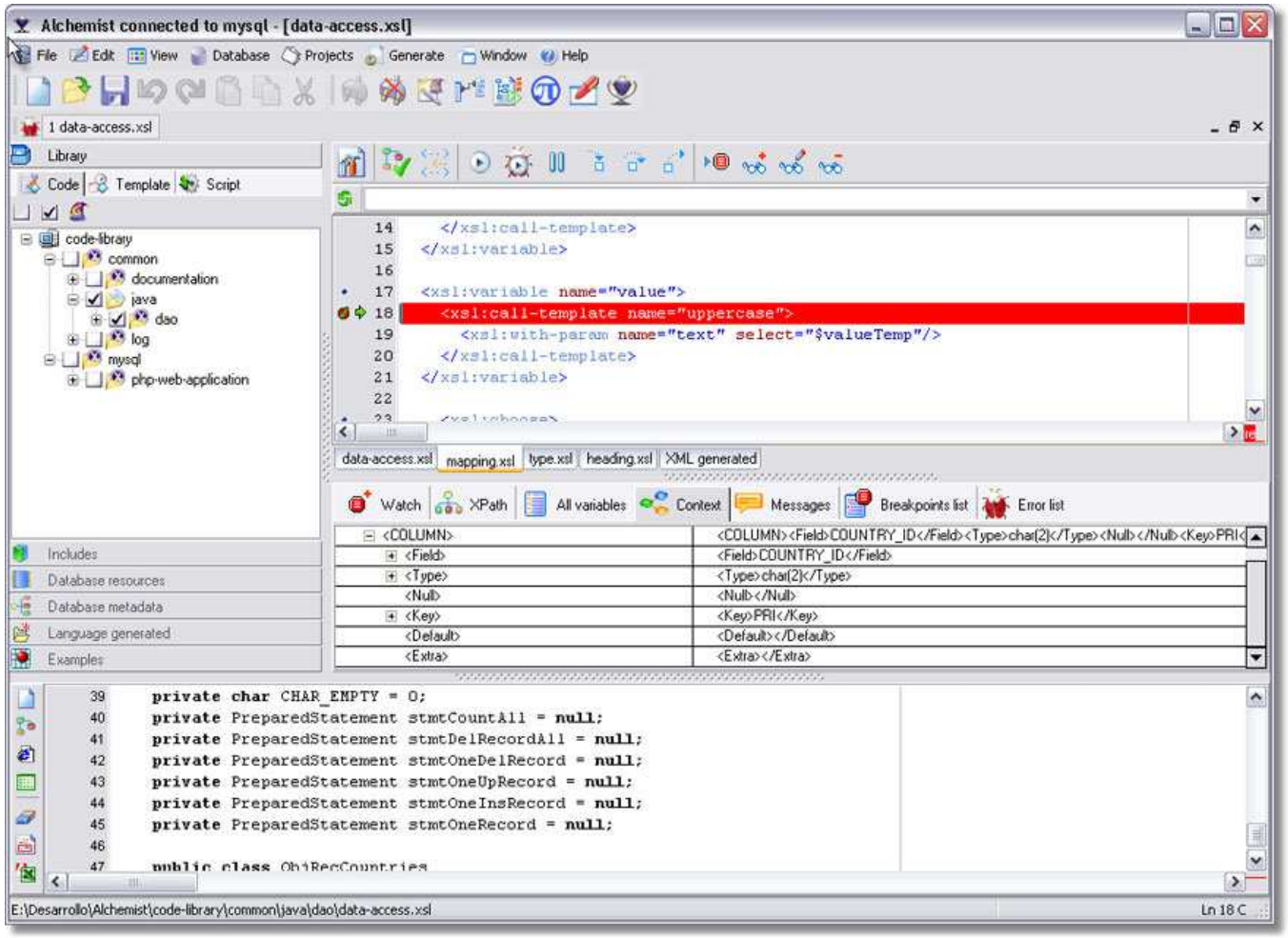
Department of Computer Science
Stevens Institute of Technology
Jan Schaumann
jschauma@stevens.edu
http://www.cs.stevens.edu/~jschauma/631/

Reminder: Final Project Requirements

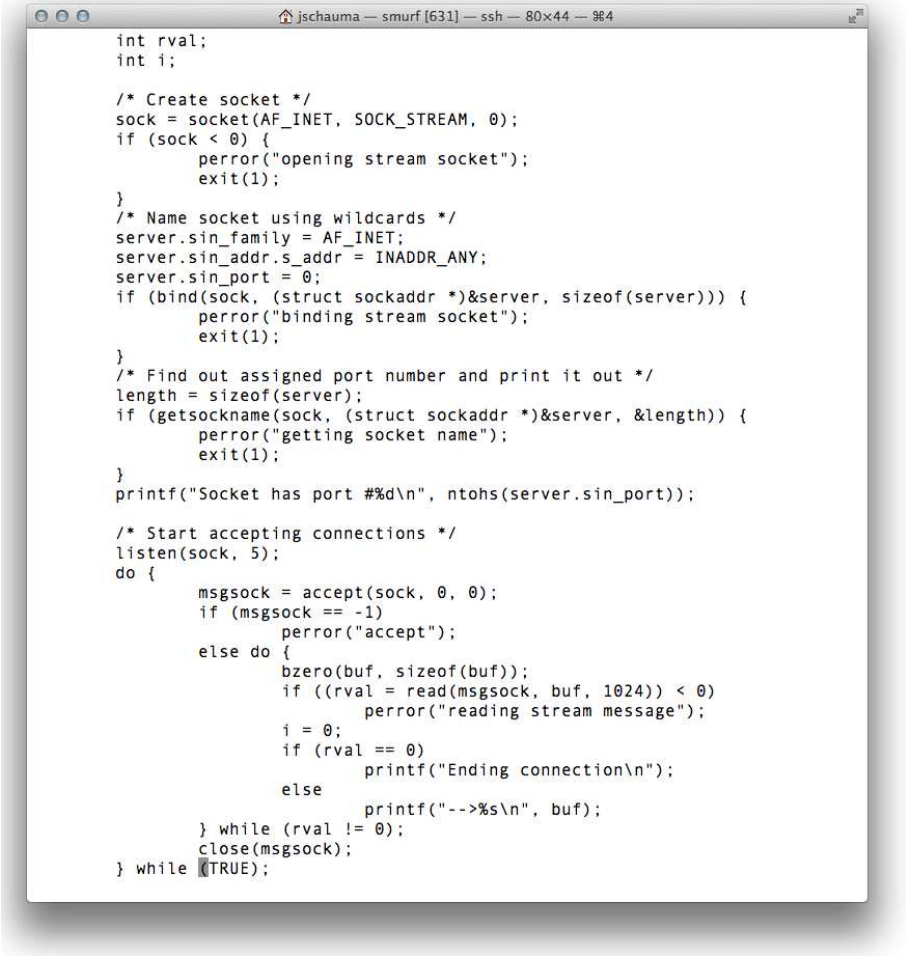
”Real World” edition

- group work:
  - 3-5 people in a team
  - all people will get the same grade on the final project
  - team-work and collaboration will be a factor in grading
- use of `git(1)` required
- code must be split across multiple files
- use of a `Makefile` required
- multi-platform and dual-stack (IPv4/IPv6) required

Software Development Tools



Software Development Tools



Software Development Tools

UNIX Userland is an IDE – essential tools that follow the paradigm of "Do one thing, and do it right" can be combined.

The most important tools are:

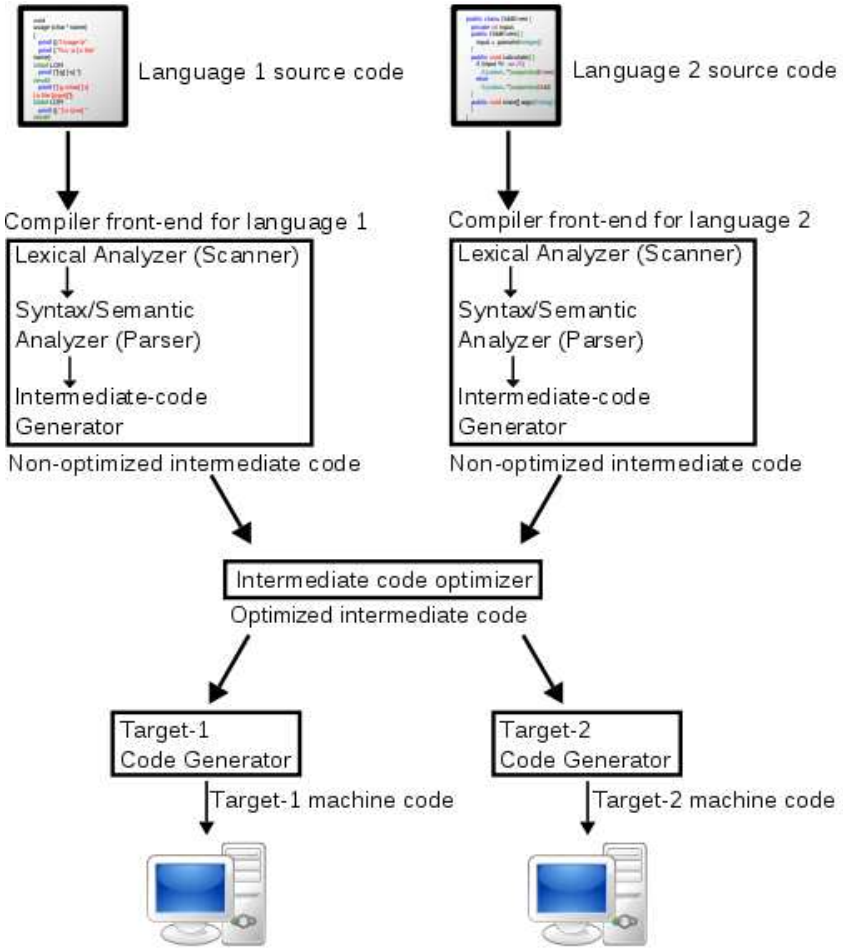
- `$EDITOR`
- the compiler toolchain
- `gdb(1)` – debugging your code
- `make(1)` – project build management, maintain program dependencies
- `diff(1)` and `patch(1)` – report and apply differences between files
- `cvs(1)`, `svn(1)`, `git(1)` etc. – distributed project management, version control

Compilers

A compiler translates *source code* from a high-level programming language into *machine code* for a given architecture by performing a number of steps:

- lexical analysis
- preprocessing
- parsing
- semantic analysis
- code generation
- code optimization

Compilers


Language 1 source code
Language 2 source code
Compiler front-end for language 1
Lexical Analyzer (Scanner)
Syntax/Semantic Analyzer (Parser)
Intermediate-code Generator
Non-optimized intermediate code
Compiler front-end for language 2
Lexical Analyzer (Scanner)
Syntax/Semantic Analyzer (Parser)
Intermediate-code Generator
Non-optimized intermediate code
Intermediate code optimizer
Optimized intermediate code
Target-1 Code Generator
Target-2 Code Generator
Target-1 machine code
Target-2 machine code

Compilers

There are many different closed- and open-source compiler chains:

- Intel C/C++ Compiler (or `icc`)
- Turbo C / Turbo C++ / C++Builder (Borland)
- Microsoft Visual C++
- ...

- Clang (a frontend to LLVM)
- GNU Compiler Collection (or `gcc`)
- Portable C Compiler (or `pcc`)
- ...

The compiler toolchain

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

Preprocessing

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

```
$ cd compilechain
$ cat hello.c
$ man cpp
$ cpp hello.c hello.i
$ file hello.i
$ man cc
$ cc -v -E hello.c > hello.i
$ more hello.i
$ cc -v -DFOOD=\"Avocado\" -E hello.c > hello.i.2
$ diff -bu hello.i hello.i.2
```

Compilation

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

```
$ more hello.i
$ cc -v -S hello.i > hello.s
$ file hello.s
$ more hello.s
```

Assembly

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

```
$ as -o hello.o hello.s
$ file hello.o
$ cc -v -c hello.s
$ objdump -d hello.o
[...]
```

Linking

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

```
$ ld hello.o
[...]
$ ld hello.o -lc
[...]
$ cc -v hello.o
[...]
$ ld -dynamic-linker /lib64/ld-linux-x86-64.so.2 \
        /usr/lib/x86_64-linux-gnu/crt1.o         \
        /usr/lib/x86_64-linux-gnu/crti.o hello.o \
        -lc /usr/lib/x86_64-linux-gnu/crtn.o
$ file a.out
$ ./a.out
```

Linking

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

```
$ cc -v -DFOOD=\"Avocado\" hello.c 2>&1 | more
```

`cc(1)` and `ld(1)`

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

Different flags can be passed to `cc(1)` to be passed through to each tool as well as to affect all tools.

```
$ cc -v -O2 -g hello.c 2>&1 | more
```

`cc(1)` and `ld(1)`

The compiler usually performs preprocessing (via `cpp(1)`), compilation (`cc(1)`), assembly (`as(1)`) and linking (`ld(1)`).

Different flags can be passed to `cc(1)` to be passed through to each tool as well as to affect all tools.

The order of the command line flags *may* play a role! Directories searched for libraries via `-L` and the resolving of undefined symbols via `-l` are examples of position sensitive flags.

```
$ cc -v main.c -L./lib2 -L./lib -lldtest 2>&1 | more
```

```
$ cc -v main.c -L./lib -L./lib2 -lldtest 2>&1 | more
```

cc(1) and ld(1)

The compiler usually performs preprocessing (via cpp(1)), compilation (cc(1)), assembly (as(1)) and linking (ld(1)).

Different flags can be passed to cc(1) to be passed through to each tool as well as to affect all tools.

The order of the command line flags *may* play a role! Directories searched for libraries via -L and the resolving of undefined symbols via -l are examples of position sensitive flags.

The behavior of the compiler toolchain may be influenced by environment variables (eg TMPDIR, SGI_ABI) and/or the compilers default configuration file (MIPSPro's /etc/compiler.defaults or gcc's specs).

```
$ cc -v hello.c
$ TMPDIR=/var/tmp cc -v hello.c
$ cc -dumpspec
```

A Debugger

gdb(1)

The purpose of a debugger such as `gdb(1)` is to allow you to see what is going on "inside" another program while it executes – or what another program was doing at the moment it crashed. `gdb` allows you to

- make your program stop on specified conditions (for example by setting *breakpoints*)
- examine what has happened, when your program has stopped (by looking at the *backtrace*, inspecting the value of certain variables)
- inspect control flow (for example by *stepping* through the program)

Other interesting things you can do:

- examine stack frames: *info frame*, *info locals*, *info args*
- examine memory: *x*
- examine assembly: *disassemble func*

gdb(1)

```
$ cc -g gdb1.c
$ ./a.out
[...]
$ gdb ./a.out
[...]
(gdb) break main
Breakpoint 1 at 0x400603: file gdb1.c, line 10.
(gdb) run
Starting program: /home/jschauma/t/gdb-examples/a.out

Breakpoint 1, main (argc=1, argv=0x7fffffffe9e8) at gdb1.c:10
10 c = fgetc(stdin);
(gdb) n
```

gdb(1)

```
$ ulimit -c unlimited
$ cc -g gdb2.c
$ ./a.out
$ gdb a.out core
bt
[...]
frame 2
li
p buf
kill
watch num
run
```

gdb(1)

```
$ cd student
$ make
$ ./a.out -lR ~djd >/dev/null
total 2701553
[...]
Memory fault
$ gdb ./a.out
 run -lR ~djd
Starting program: /home/jschauma/t/student/a.out -lR ~djd
total 2701553
[...]

Program received signal SIGSEGV, Segmentation fault.
0x000000000040214a in print_entries (entryvect=0x6050a0, options=0x605010) at print.c
221                     printf("%10s ", grpentry->gr_name);
```

gdb(1)

```
(gdb) bt
#0  0x000000000040214a in print_entries (entryvect=0x6050a0, options=0x605010) at pri
#1  0x0000000000401ad5 in list_dir_contents (path=0x7fffffffec59 "/home/djd", ftsopts
#2  0x000000000040292e in main (argc=1, argv=0x7fffffffe9e8) at ls.c:57
(gdb) li
216   } else {
217       pwentry = getpwuid(myentryp->fts_statp->st_uid);
218       printf("%8s ", pwentry->pw_name);
219
220       grpentry = getgrgid(myentryp->fts_statp->st_gid);
221       printf("%10s ", grpentry->gr_name);
222   }
223
224   // displaying the size
225   size = (long long)(myentryp->fts_statp->st_size);
(gdb) p grpentry
$1 = (struct group *) 0x0
```

make(1)

`make(1)`

`make(1)` is a command generator and build utility. Using a description file (usually *Makefile*) it creates a sequence of commands for execution by the shell.

- used to sort out dependency relations among files

make(1)

make(1) is a command generator and build utility. Using a description file (usually *Makefile*) it creates a sequence of commands for execution by the shell.

- used to sort out dependency relations among files
- avoids having to rebuild the entire project after modification of a single source file

`make(1)`

`make(1)` is a command generator and build utility. Using a description file (usually *Makefile*) it creates a sequence of commands for execution by the shell.

- used to sort out dependency relations among files
- avoids having to rebuild the entire project after modification of a single source file
- performs *selective* rebuilds following a *dependency graph*

`make(1)`

`make(1)` is a command generator and build utility. Using a description file (usually *Makefile*) it creates a sequence of commands for execution by the shell.

- used to sort out dependency relations among files
- avoids having to rebuild the entire project after modification of a single source file
- performs *selective* rebuilds following a *dependency graph*
- allows simplification of rules through use of *macros* and *suffixes*, some of which are internally defined

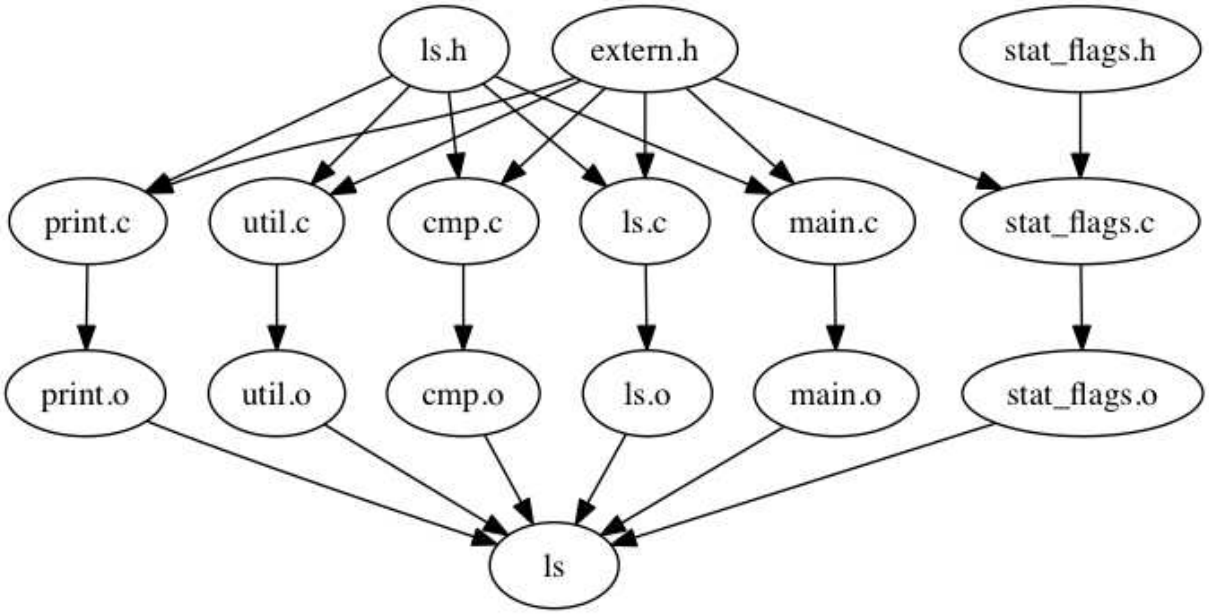
`make(1)`

`make(1)` is a command generator and build utility. Using a description file (usually *Makefile*) it creates a sequence of commands for execution by the shell.

- used to sort out dependency relations among files
- avoids having to rebuild the entire project after modification of a single source file
- performs *selective* rebuilds following a *dependency graph*
- allows simplification of rules through use of *macros* and *suffixes*, some of which are internally defined
- different versions of `make(1)` (BSD make, GNU make, Sys V make, ...) may differ (among other things) in
  - variable assignment and expansion/substitution
  - including other files
  - flow control (for-loops, conditionals etc.)

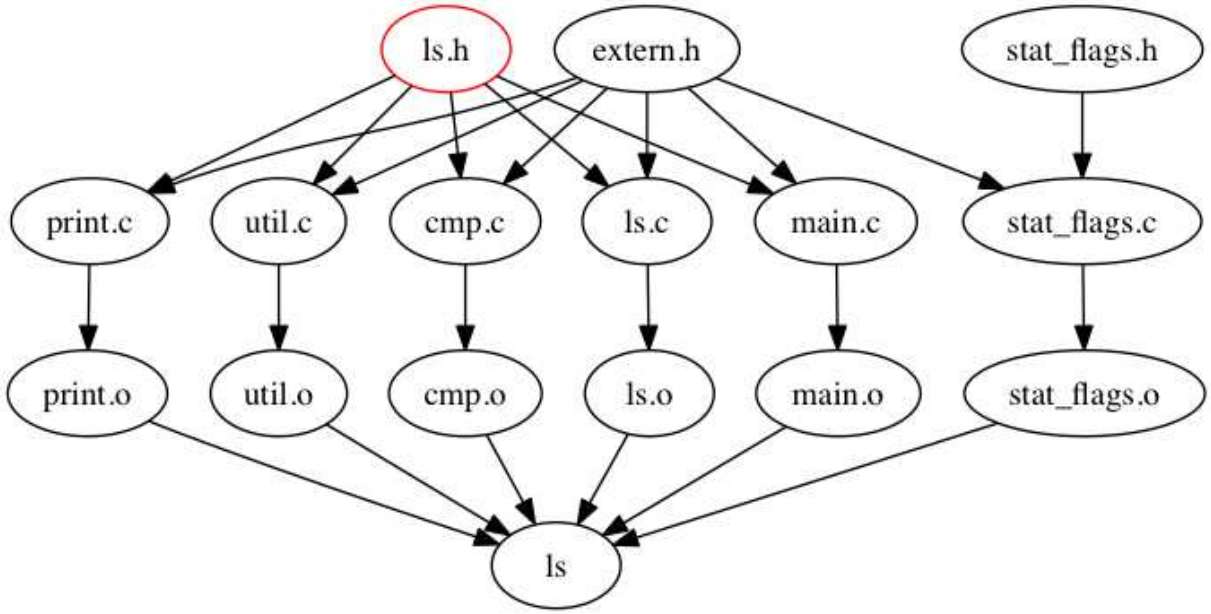
make(1)

```
$ cd make-examples
$ ls *.[ch]
cmp.c       ls.c        main.c        stat_flags.c        util.c
extern.h    ls.h        print.c       stat_flags.h
```



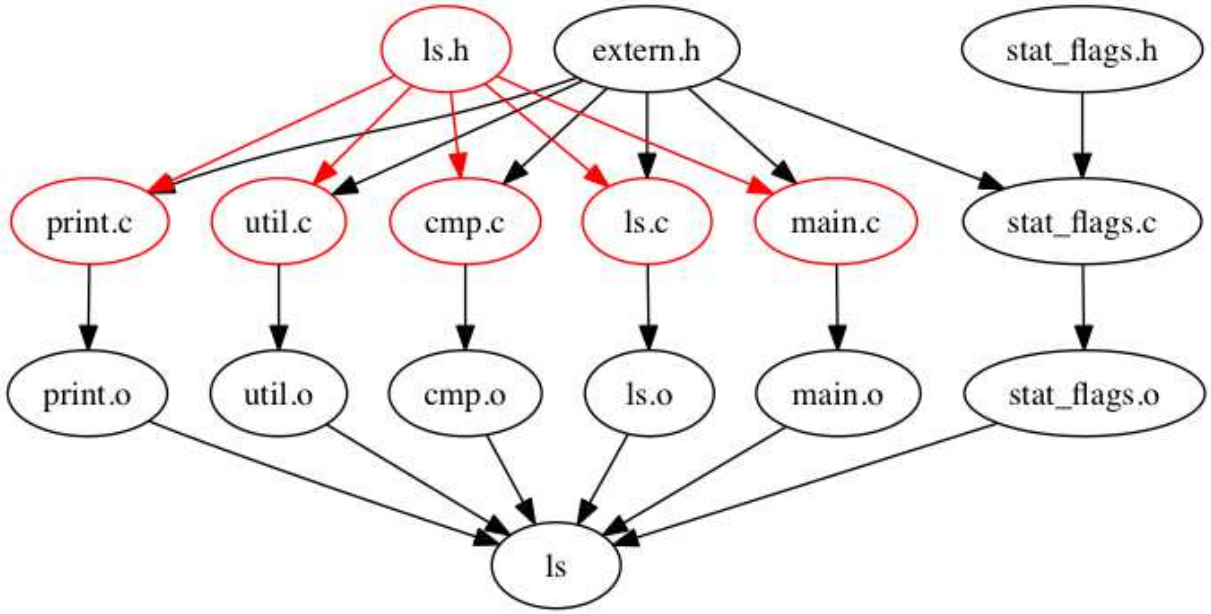
make(1)

```
$ cd make-examples
$ ls *.[ch]
cmp.c       ls.c        main.c        stat_flags.c        util.c
extern.h    ls.h        print.c       stat_flags.h
```



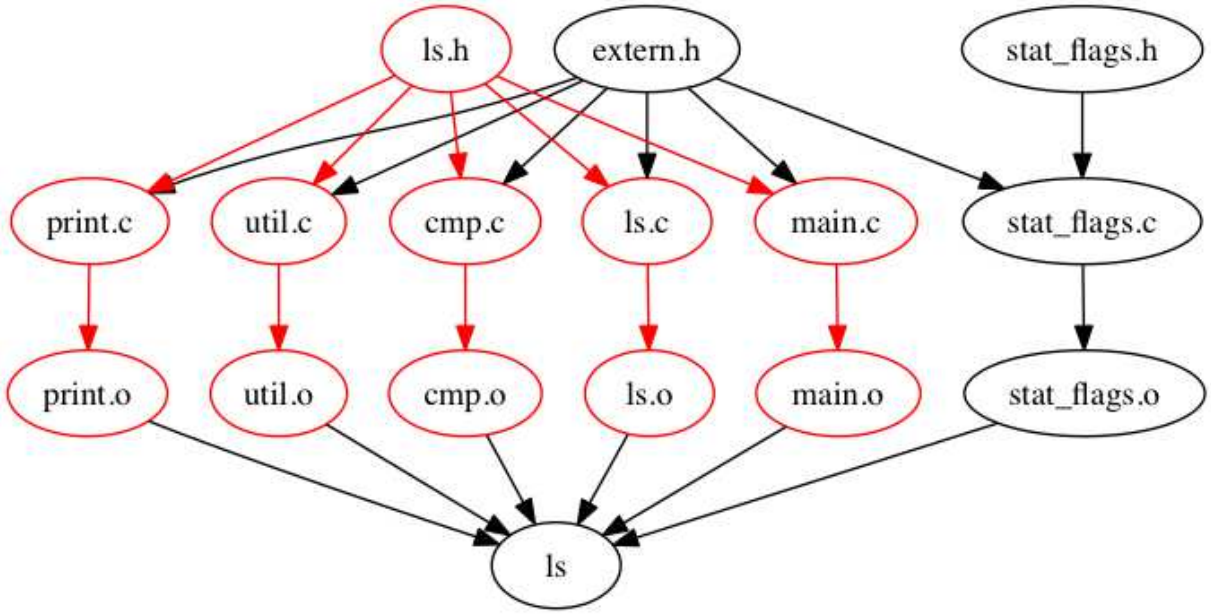
make(1)

```
$ cd make-examples
$ ls *.[ch]
cmp.c       ls.c        main.c        stat_flags.c        util.c
extern.h    ls.h        print.c       stat_flags.h
```



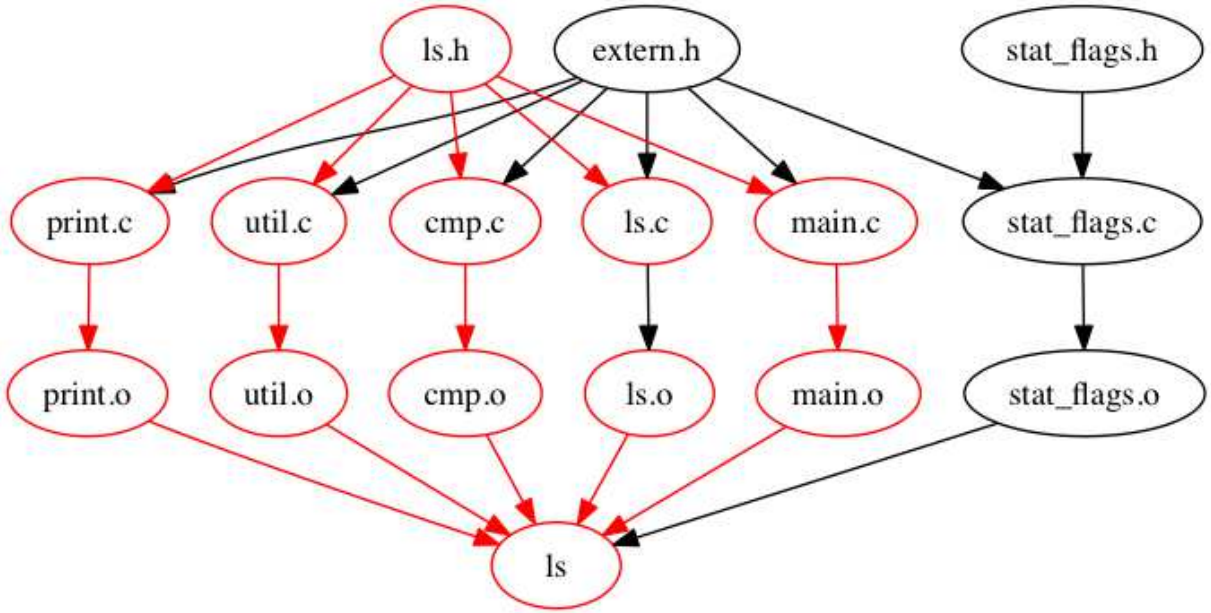
make(1)

```
$ cd make-examples
$ ls *.[ch]
cmp.c       ls.c        main.c        stat_flags.c        util.c
extern.h    ls.h        print.c       stat_flags.h
```



make(1)

```
$ cd make-examples
$ ls *.[ch]
cmp.c       ls.c        main.c        stat_flags.c        util.c
extern.h    ls.h        print.c       stat_flags.h
```



make(1)

```
$ make -f Makefile.1
cc -c cmp.c
cc -c ls.c
cc -c main.c
cc -c print.c
cc -c stat_flags.c
cc -c util.c
cc cmp.o ls.o main.o print.o stat_flags.o util.o -o ls
$ touch ls.h
$ make -f Makefile.1
cc -c cmp.c
cc -c ls.c
cc -c main.c
cc -c print.c
cc -c util.c
cc cmp.o ls.o main.o print.o stat_flags.o util.o -o ls
$
```

make(1)

```
$ make -f Makefile.2
cc    -c -o cmp.o cmp.c
cc    -c -o ls.o ls.c
[...]
cc    -c -o util.o util.c
ls depends on cmp.o ls.o main.o print.o stat_flags.o util.o
cc  cmp.o ls.o main.o print.o stat_flags.o util.o -o ls
$ bmake -f Makefile.2
gcc -pipe -O2  -c cmp.c
gcc -pipe -O2  -c ls.c
[...]
gcc -pipe -O2  -c util.c
ls depends on cmp.o ls.o main.o print.o stat_flags.o util.o
gcc  cmp.o ls.o main.o print.o stat_flags.o util.o -o ls
$
```

make(1)

```
$ make -f Makefile.3
cc -c cmp.c -o cmp.b
cc -c ls.c -o ls.b
cc -c main.c -o main.b
cc -c print.c -o print.b
cc -Wall -g -c stat_flags.c -o stat_flags.bar
cc -Wall -g -c util.c -o util.bar
ls depends on cmp.b ls.b main.b print.b stat_flags.bar util.bar
cc  cmp.b ls.b main.b print.b stat_flags.bar util.bar -o ls
$
```

Priority of Macro Assignments for `make(1)`

1. Internal (default) definitions of `make(1)`
2. Current shell environment variables. This includes macros that you enter on the *make* command line itself.
3. Macro definitions in *Makefile*.
4. Macros entered on the `make(1)` command line, if they follow the *make* command itself.

make(1)

```
$ bmake -f Makefile.4
[...]
$ bmake -f Makefile.2
[...]
$ CFLAGS=-Werror bmake -f Makefile.2
[...]
$ CFLAGS=-Werror bmake -f Makefile.4
[...]
$ bmake CFLAGS=-Werror -f Makefile.2
[...]
$ bmake CFLAGS=-Werror -f Makefile.4
[...]
```

Ed is the standard text editor.

```
$ ed
?
help
?
quit
?
exit
?
bye
?
eat flaming death
?
^C
?
^D
?
```

Ed is the standard text editor.

```
$ ed
a
ed is the standard Unix text editor.
This is line number two.
.
2i

.
%l
3s/two/three/
w foo
q
$ cat foo
```

`diff(1)` and `patch(1)`

`diff(1)`:

- compares files line by line
- output may be used to automatically edit a file
- can produce human “readable” output as well as diff entire directory structures
- output called a *patch*

`diff(1)` and `patch(1)`

`patch(1)`:

- applies a `diff(1)` file (aka *patch*) to an original
- may back up original file
- may guess correct format
- ignores leading or trailing “garbage”
- allows for reversing the patch
- may even correct context line numbers

diff(1) and patch(1)

```
$ diff Makefile.2 Makefile.4
8c8
< #CFLAGS= -Wall -g
---
> CFLAGS= -Wall -g
$ cp Makefile.2 /tmp
$ ( diff -e Makefile.2 Makefile.4; echo w; ) | ed Makefile.2
$ diff Makefile.[24]
$ mv /tmp/Makefile.2 .
$ diff -c Makefile.[24]
$ diff -u Makefile.[24] > /tmp/patch
$ patch </tmp/patch
$ diff Makefile.[24]
```

Revision Control

Version control systems allow you to

- collaborate with others
- simultaneously work on a code base
- keep old versions of files
- keep a log of the who, when, what, and why of any changes
- perform release engineering by creating *branches*

Revision Control

- Source Code Control System (*SSCS*) begat the Revision Control System (*RCS*).
- RCS operates on a single file; still in use for misc. OS config files
- the Concurrent Versions System (*CVS*) introduces a client-server architecture, control of hierarchies
- *Subversion* provides atomic commits, renaming, cheap branching etc.
- *Git*, *Mercurial* etc. implement a *distributed* approach (ie peer-to-peer versus client-server), adding other features (cryptographic authentication of history, ...)

Revision Control

Examples:

`http://cvsweb.netbsd.org/bsdweb.cgi/src/bin/ls/`

`http://svnweb.freebsd.org/base/stable/9/bin/ls/`

`http://git.savannah.gnu.org/cgit/coreutils.git/log/`

`http://cvsweb.netbsd.org/bsdweb.cgi/src/share/misc/bsd-family-tree?rev=HEAD`

Revision Control: Branching

Different strategies:

- trunk / master is fragile
  - *trunk* is work in progress, may not even compile
  - all work happens in *trunk*
  - releases are tagged on *trunk*, then branched
- trunk / master is stable
  - *master* is always stable
  - all work is done in branches (feature or bugfix)
  - feature branches are deleted after merge
  - releases are made automatically from master

You may combine these as *release branching* / *feature branching* / *task branching*.

Commit Messages

Commit messages are like comments: often useless and misleading, but critical in understanding human thinking behind the code.

Commit messages are full sentences in correct and properly formatted English.

Commit messages briefly summarize the *what*, but provide important historical context as to the *how* and, more importantly, *why*.

Commit messages SHOULD reference and integrate with ticket tracking systems.

See also:

- `http://is.gd/Wd1LhA`
- `http://is.gd/CUtwhA`
- `http://is.gd/rPQj5E`

Links

Revision Control:

`http://cvsbook.red-bean.com/cvsbook.html`

`http://svnbook.red-bean.com/`

`http://git-scm.com/`

GDB:

`http://sources.redhat.com/gdb/current/onlinedocs/gdb_toc.html`

`http://heather.cs.ucdavis.edu/~matloff/UnixAndC/CLanguage/Debug.html`

`http://www.unknownroad.com/rtfm/gdbtut/gdbtoc.html`