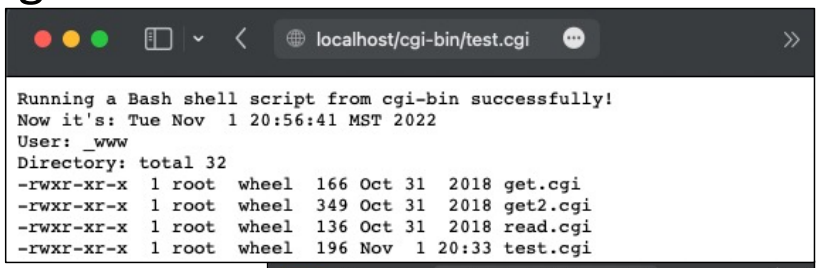
Lecture 20

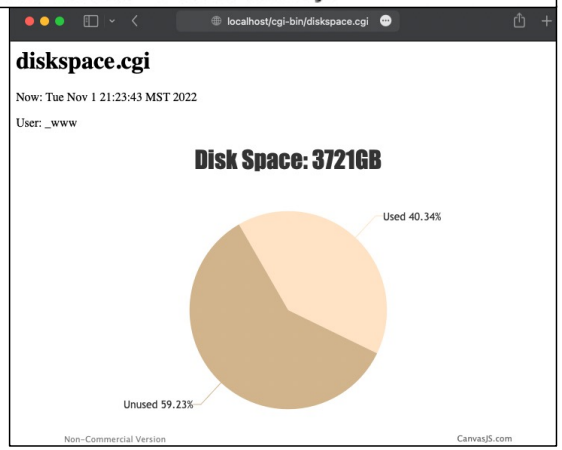
408/508 *Computational Techniques for Linguists*

Today's Topics

- Last time we discussed running programs on the webserver itself.
  - two examples: one `text/plain`,
  - one `text/html`
  - both from the default `cgi-bin` directory

1. Today, running a program from inside your home directory



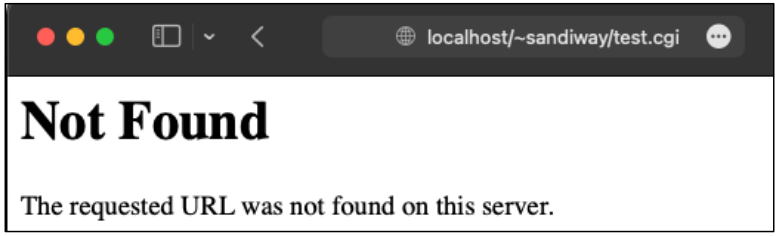


Apache2 `cgi-bin` directories

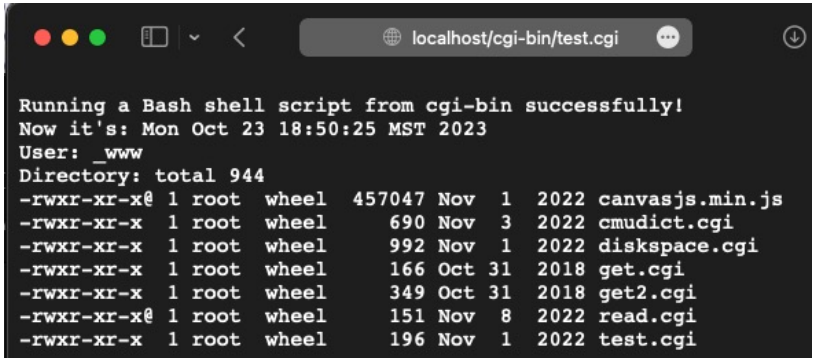
- Default `cgi-bin` directory
  - `http://localhost/cgi-bin/`*`test.cgi`*
  - macOS: `/Library/WebServer/CGI-Executables/` Ubuntu: `/usr/lib/cgi-bin/`
- User directories
  - `http://localhost/`*`~username`*`/`*`test.cgi`*
  - macOS: `~/Sites` Ubuntu: `~/public_html`
  - *Users are not permitted access to the default cgi-bin directory*
  - We can permit cgi-bin scripts in **a designated user directory**

Apache2 Webserver on macOS

- cgi-bin not working from user directory!



- cgi-bin working from default directory!



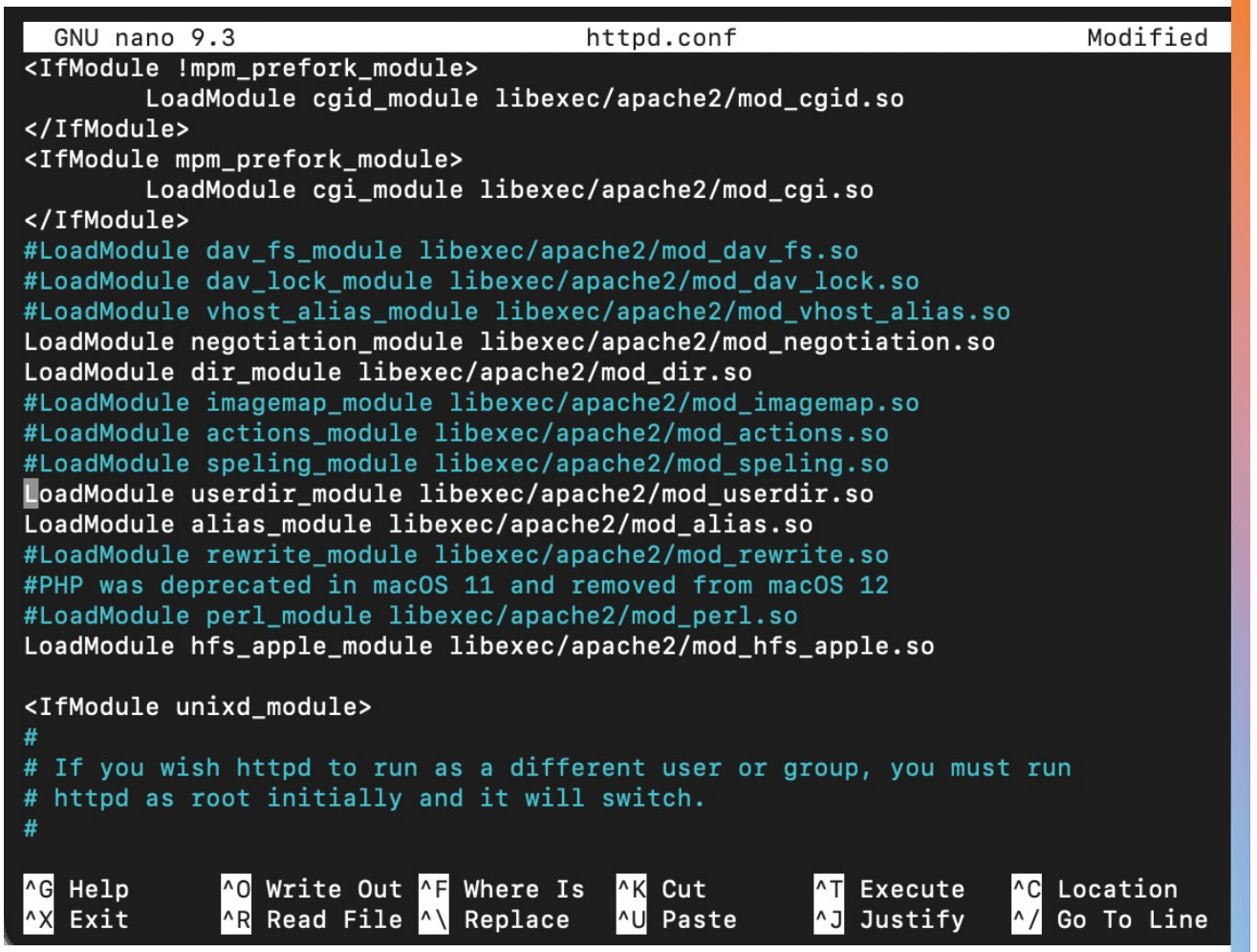
- Note:
  - http://localhost/~user/test.cgi should map to ~user/Sites/test.cgi
  - By default, it's turned off (*for security*)
  - Even if httpd is running … check with `ps –ax | grep httpd`

Apache2 Webserver on macOS

- To run programs in `~/Sites`,
- i.e. outside of `/Library/WebServer/CGI-Executables`
  - modify the Apache2 httpd configuration file:
  - `sudo nano /etc/apache2/httpd.conf`
- Somewhat madding:
  - after macOS updates: *it nukes my changes to* `httpd.conf` (*and other files*)!
- After modifying:
  - `sudo apachectl stop`
  - `sudo apachectl start`

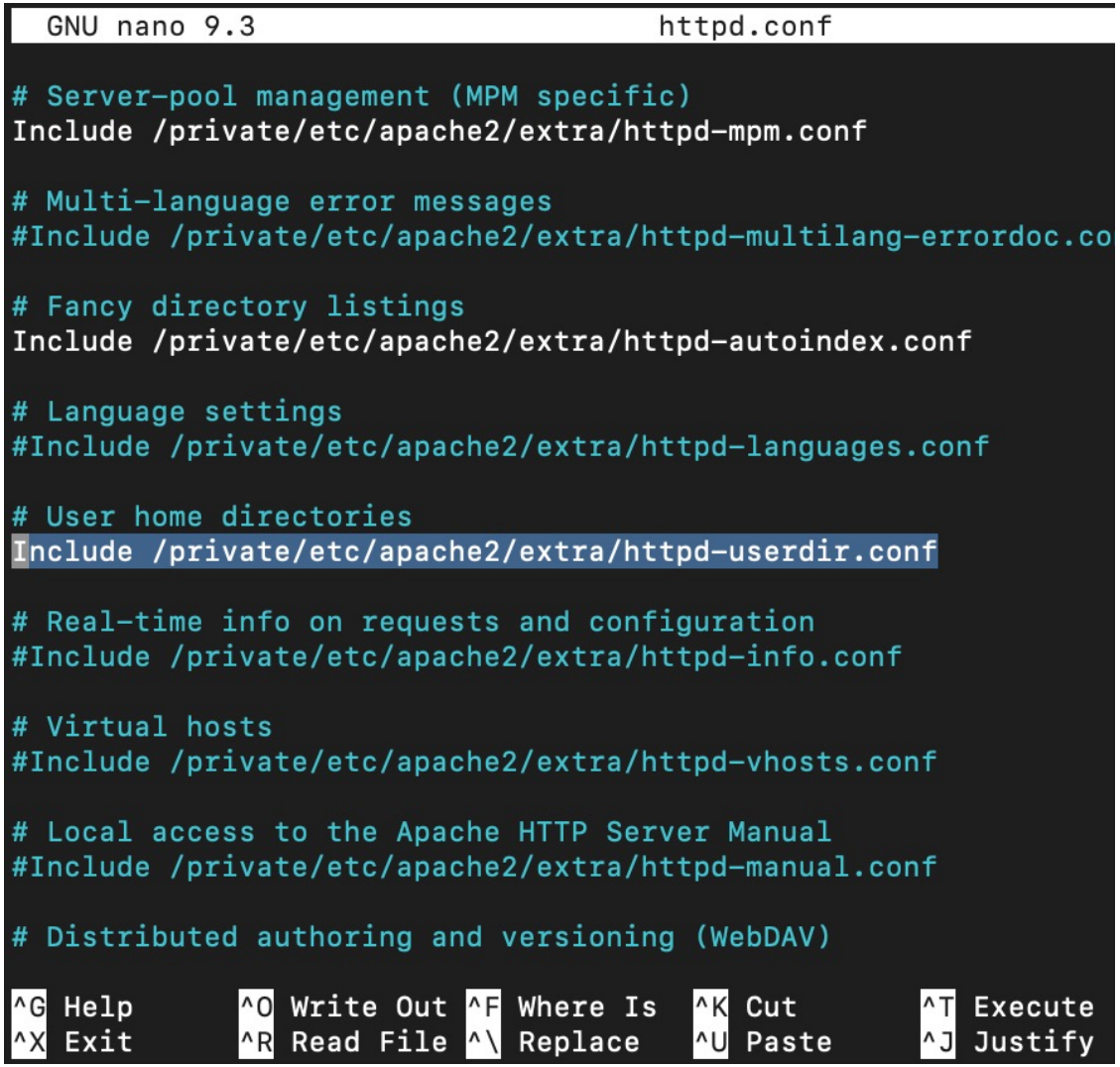
Apache2 Webserver on macOS

- In `/etc/apache2/httpd.conf`
- *uncomment `userdir_module` line!*



Apache2 Webserver on macOS

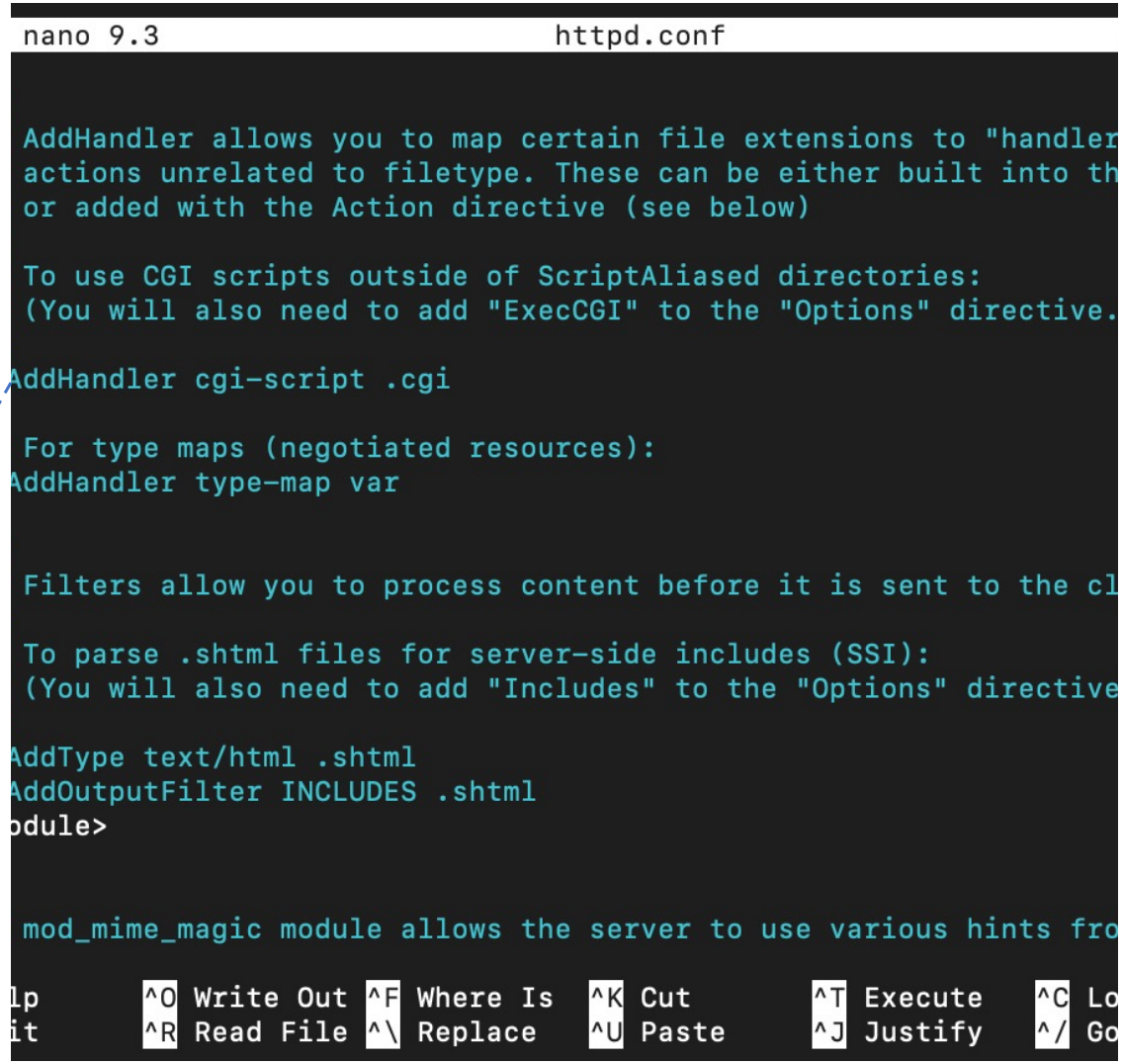
- In `/etc/apache2/httpd.conf`
- make sure `httpd-userdir.conf` line is not commented out



Apache2 Webserver on macOS

- In `/etc/apache2/httpd.conf`
- uncomment line `AddHandler cgi-script .cgi`

- *this invokes the cgi-script handler for all files of type .cgi*
  - therefore, all our scripts should have extension .cgi



Apache2 Webserver on macOS

```
apache2$ diff httpd.conf httpd.conf.sonoma
171c171
< LoadModule cgid_module libexec/apache2/mod_cgid.so
---
> #LoadModule cgid_module libexec/apache2/mod_cgid.so
174c174
< LoadModule cgi_module libexec/apache2/mod_cgi.so
---
> #LoadModule cgi_module libexec/apache2/mod_cgi.so
184c184
< LoadModule userdir_module libexec/apache2/mod_userdir.so
---
> #LoadModule userdir_module libexec/apache2/mod_userdir.so
447c447
<     AddHandler cgi-script .cgi
---
>     #AddHandler cgi-script .cgi
521c521
< Include /private/etc/apache2/extra/httpd-userdir.conf
---
> #Include /private/etc/apache2/extra/httpd-userdir.conf
556a557
>
```

Apache2 Webserver on macOS

- `sudo nano /etc/apache2/extra/httpd-userdir.conf`

```
1# Settings for user home directories¶
2#¶
3# Required module: mod_authz_core, mod_authz_host, mod_userdir¶
4¶
5#¶
6# UserDir: The name of the directory that is appended onto a user's home¶
7# directory if a ~user request is received.  Note that you must also set¶
8# the default access control for these directories, as in the example below.¶
9#¶
10UserDir Sites¶
11¶
12#¶
13# Control access to UserDir directories.  The following is an example¶
14# for a site where these directories are restricted to read-only.¶
15#¶
16#Include /private/etc/apache2/users/*.conf¶
17<IfModule bonjour_module>¶
18       RegisterUserSite customized-users¶
19</IfModule>¶
```



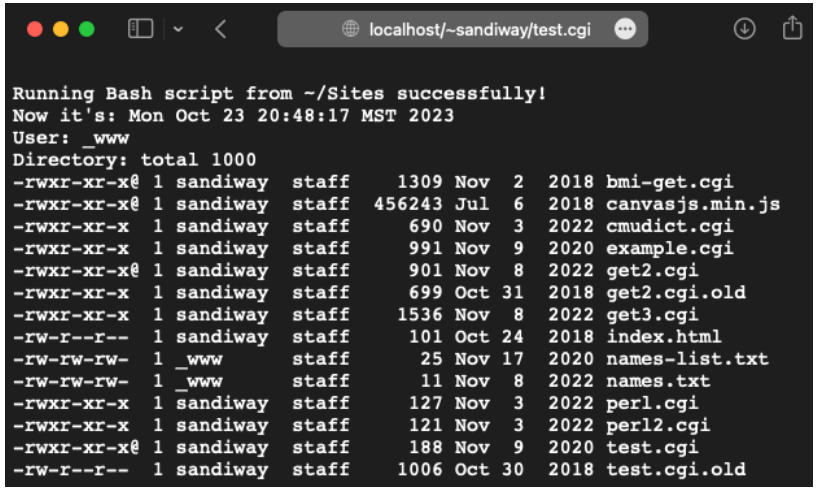
Apache2 Webserver on macOS

- `sudo nano /etc/apache2/users/username.conf`

```
1<Directory "/Users/sandiway/Sites/">¶
2        AllowOverride All¶
3        Options Indexes MultiViews FollowSymLinks ExecCGI¶
4        Require all granted¶
5</Directory>¶
```

  - *create this XML file for your username if it doesn't exist*
    - *must have ExecCGI there!*

Apache2 Webserver on macOS



check permissions must be executable `-rwxr-xr-x`

- Note user _www!
- ~*username*/Sites/test.cgi

```
1 #!/bin/bash
2 echo "Content-Type: text/plain"
3 echo
4 echo "Running Bash script from ~/Sites successfully!"
5 echo -n "Now it's: "
6 date
7 echo -n "User: "
8 whoami
9 echo -n "Directory: "
10 ls -l
11 exit 0
```

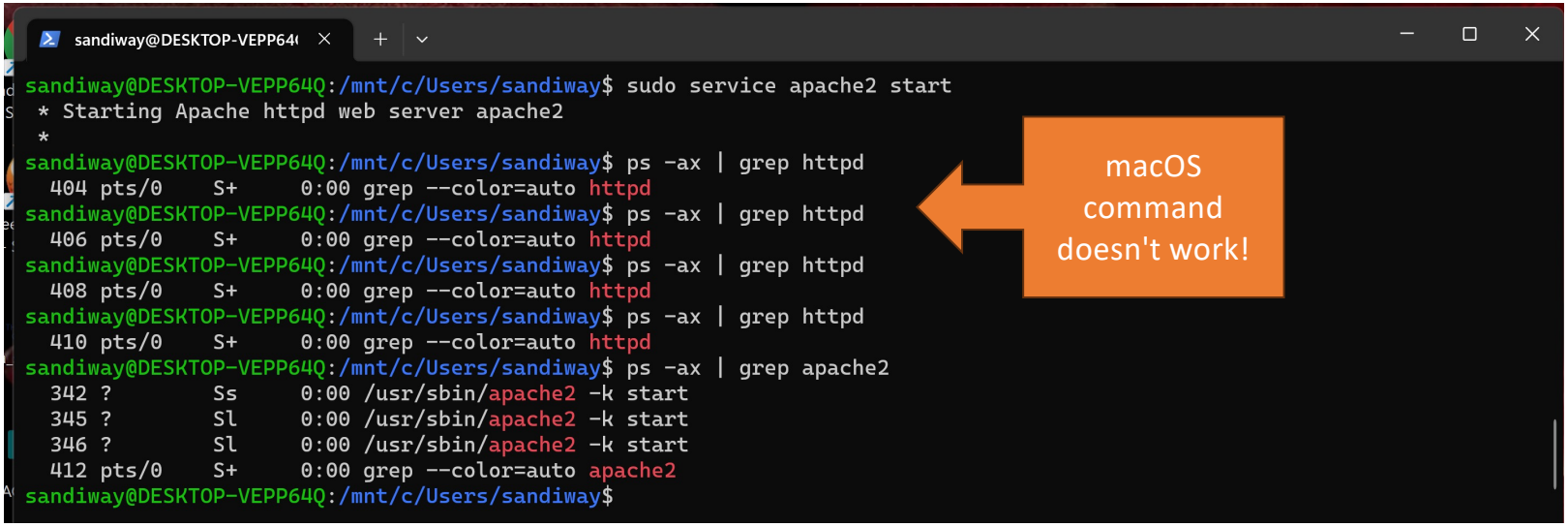
Same as test.cgi on course website except for echo here

Note the format of the first 3 lines:

1. `#!/bin/bash` (*use bash to run this*)
2. `Content-type: text/html` (*echo*)
3. (blank line) (*echo*)

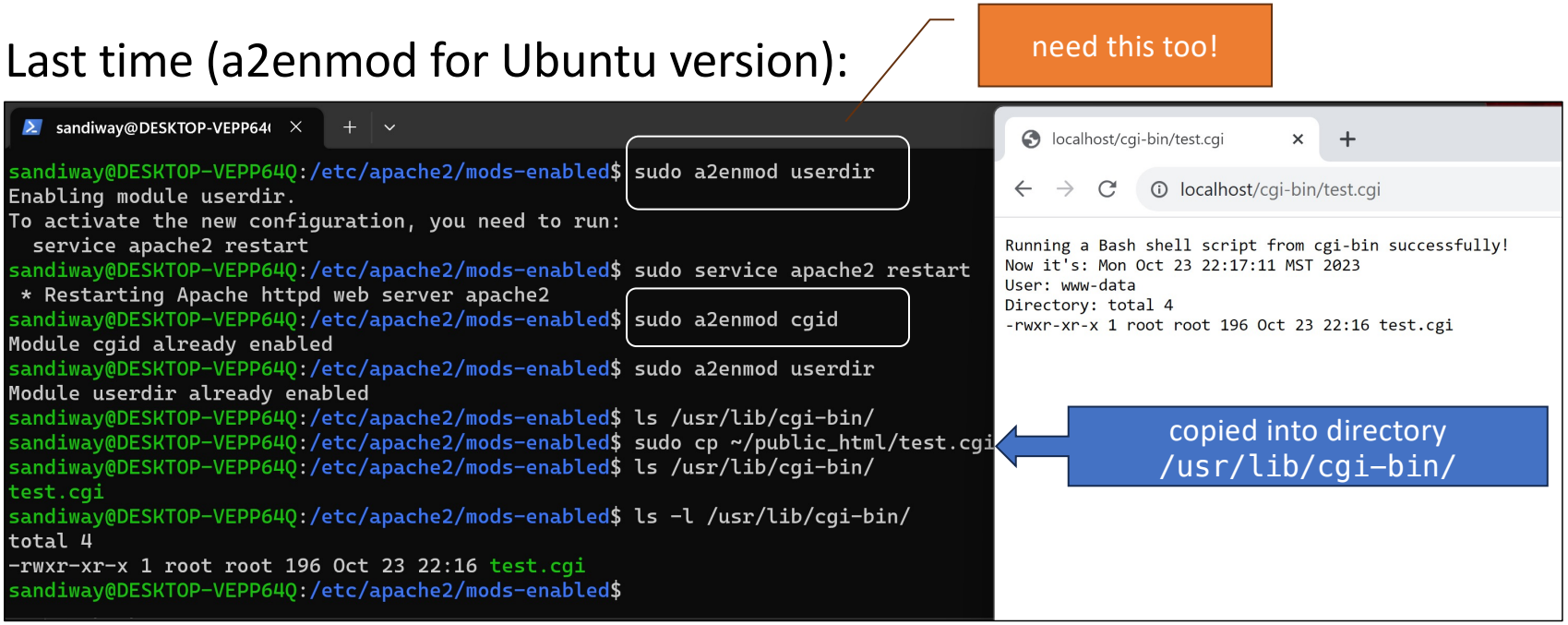
Apache Webserver on Ubuntu

- `Windows Powershell: wsl (or ubuntu)`
- `sudo service apache2 start`
- To see if apache2 is running `ps –ax | grep apache2`:



Apache Webserver on Ubuntu

- Last time (a2enmod for Ubuntu version):



Apache Webserver on Ubuntu

- https://manpages.ubuntu.com/manpages/jammy/en/man8/a2enmod.8.html

NAME

a2enmod, a2dismod – enable or disable an apache2 module

SYNOPSIS

a2enmod [-q|--quiet] [-m|--maintmode] [module]

a2dismod [-q|--quiet] [-f|--force] [-m|--maintmode] [-p|--purge] [module]

DESCRIPTION

This manual page documents briefly the **a2enmod** and **a2dismod** commands.

a2enmod is a script that enables the specified module within the **apache2** configuration. It does this by creating symlinks within **/etc/apache2/mods-enabled**. Likewise, **a2dismod** disables a module by removing those symlinks. It is not an error to enable a module which is already enabled, or to disable one which is already disabled.

Note that many modules have, in addition to a .load file, an associated .conf file. Enabling the module puts the configuration directives in the .conf file as directives into the main server context of **apache2**.

Apache2 Webserver on Ubuntu

- By default, `cgi-bin` is not enabled for `~/public_html` (macOS `~/Sites`)
- https://httpd.apache.org/docs/2.4/howto/cgi.html
- **CGI outside of ScriptAlias directories**
- CGI programs are often restricted to ScriptAlias'ed directories for security reasons. In this way, administrators can tightly control who is allowed to use CGI programs. However, if the proper security precautions are taken, there is no reason why CGI programs cannot be run from arbitrary directories.
- For example, you may wish to let users have web content in their home directories with the UserDir directive.



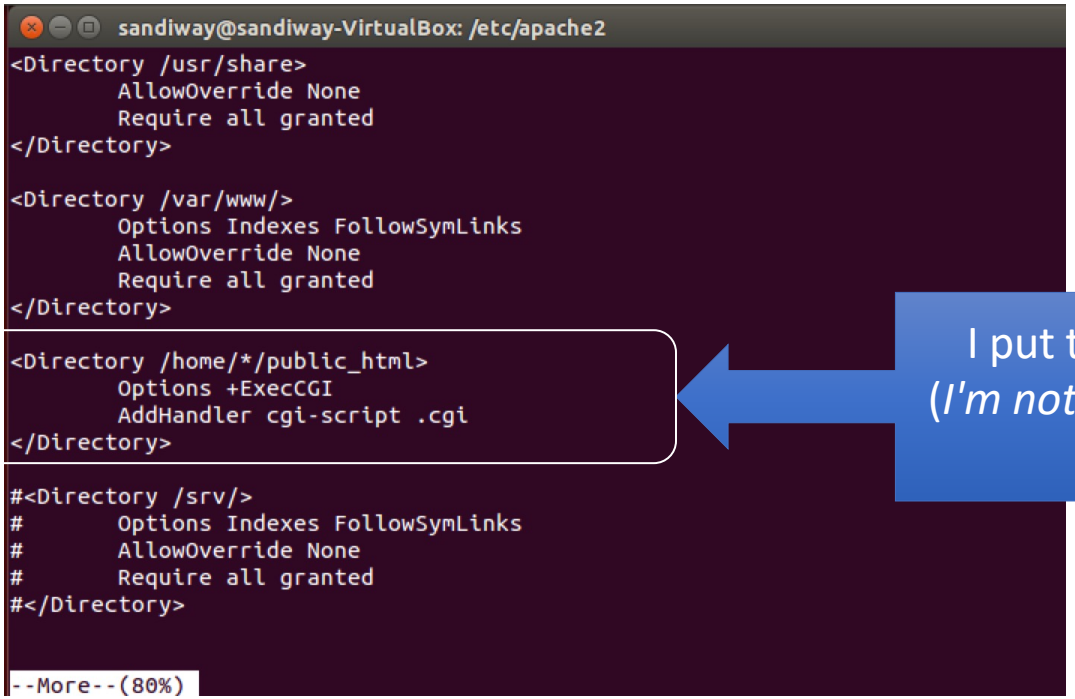
`mkdir ~/public_html` (if directory doesn't exist)
`~` should be `/home/`*`username`* (`cd ~` *should take you there*)
`/mnt/c/Users/`*`username`* (`default on WSL` *is not your Ubuntu home!*)

Apache Webserver on Ubuntu

- From https://httpd.apache.org/docs/current/howto/cgi.html
  - add these lines to `/etc/apache2/apache.conf`
    - `<Directory "/home/*/public_html">`
    - `    Options +ExecCGI`
    - `    AddHandler cgi-script .cgi`
    - `</Directory>`
  - and restart apache2 after the modifications:
    - `sudo service apache2 restart`

Apache Webserver on Ubuntu

- `sudo nano /etc/apache2/apache2.conf`



I put the declarations here
(*I'm not sure it's the best place though…*)

Apache Webserver on Ubuntu

- `sudo nano /etc/apache2/mods-available/userdir.conf`

```
<IfModule mod_userdir.c>
        UserDir public_html
        UserDir disabled root

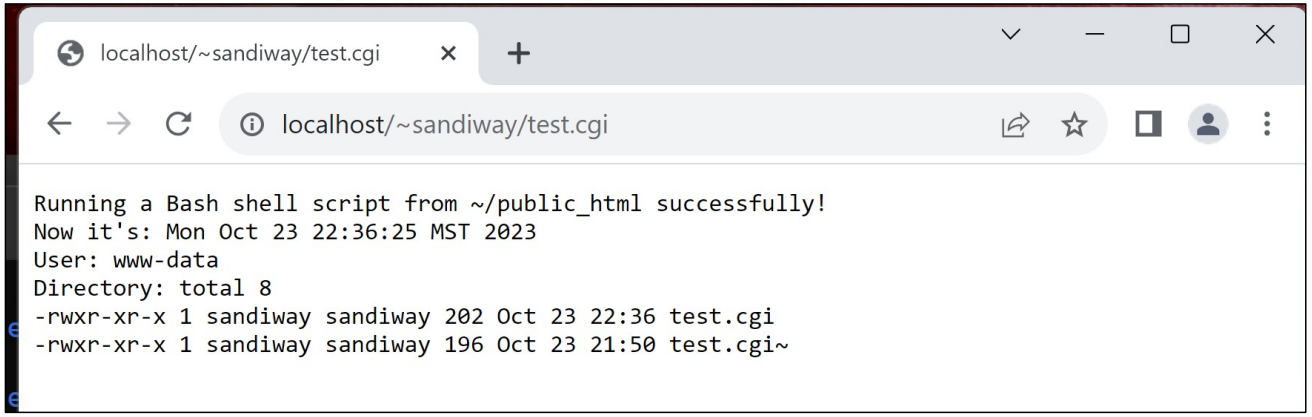
        <Directory /home/*/public_html>
                AllowOverride FileInfo AuthConfig Limit Indexes
                Options MultiViews Indexes SymLinksIfOwnerMatch IncludesNoExec
                <Limit GET POST OPTIONS>
                        Require all granted
                </Limit>
                <LimitExcept GET POST OPTIONS>
                        Require all denied
                </LimitExcept>
        </Directory>
</IfModule>

# vim: syntax=apache ts=4 sw=4 sts=4 sr noet
```

not sure this exactly is necessary, but it works!

Apache Webserver on Ubuntu

- ~*sandiway*/public_html/test.cgi



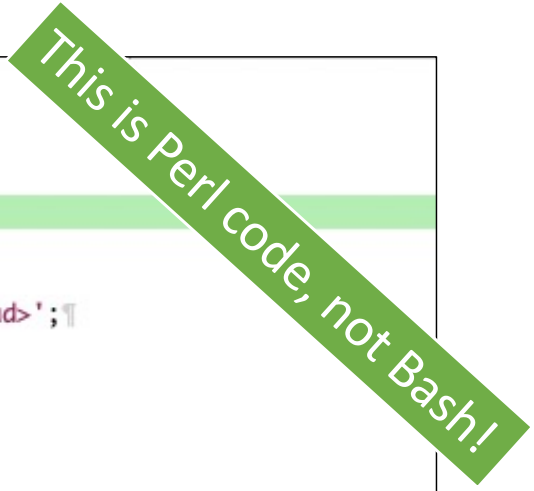
Worked example: `CMUDict`

- Suppose:
  - we want to serve the *CMU Pronouncing Dictionary* (CMUDict) – *see next slide*
  - assume it's already installed in Perl (*a programming language*)
- **How?**
  1. assume script is executable (`chmod 755`)
  2. 1st line of script, e.g. `#!/usr/bin/perl` tells the shell what program to use
  3. Then remaining lines of the script will be Perl code
  4. `which perl` gives the absolute PATH for Perl as, e.g., `/usr/bin/perl`
  - *same for Python and any other programming languages etc.*
- **Debugging:**
  - `tail /var/log/apache2/error_log` (`error.log`)
  - `tail /var/log/apache2/access_log` (`access.log`)

Worked example: `CMUDict`

- `cmudict.cgi` on course website

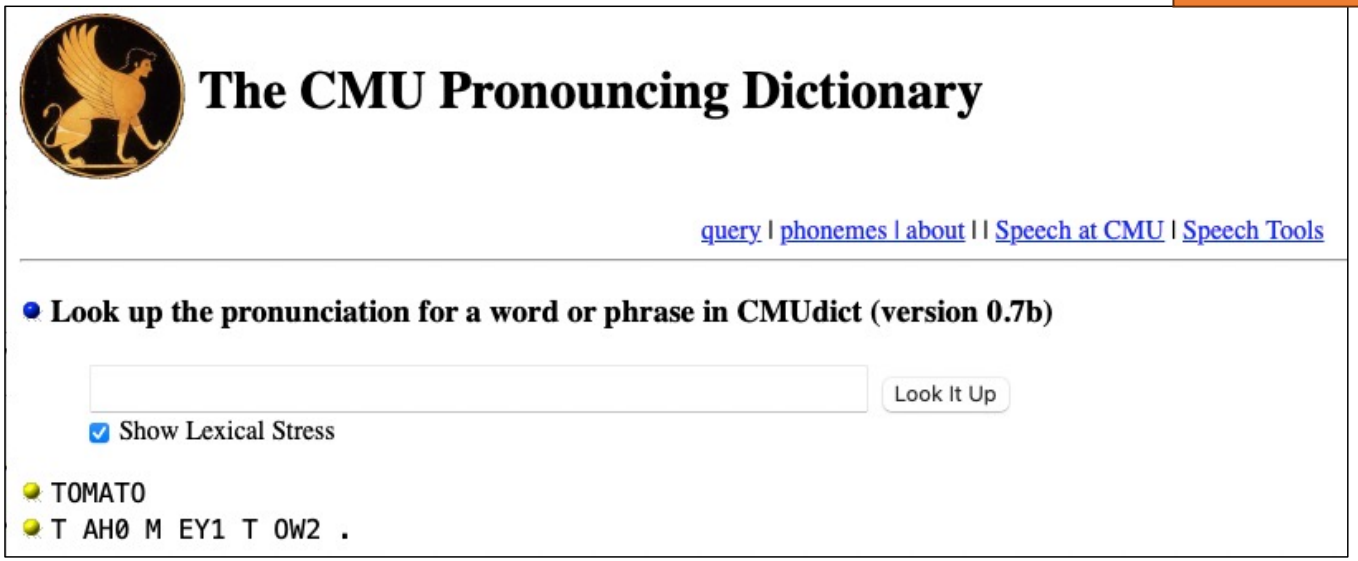
```
1#!/usr/bin/perl
2use Lingua::EN::CMUDict;
3my $obj = new Lingua::EN::CMUDict;
4my $string = $ENV{QUERY_STRING};
5my $word =  $string ? $string : $ARGV[0]; # remove word=
6my $n = $obj->number_of_syllables($word);
7print "Content-type: text/html; charset=utf-8\n\n";
8print '<html><head><style>div {font-size: x-large}</style></head>';
9print '<body><h1>Using CMUDict from Perl</h1><div>';
10if ($n) {
11  print "<em>$word</em> has $n syllable(s).</div>";
12  my $pron = $obj->get_word($word);
13  print "$pron<br>";
14} else {
15  print "<em>$word</em> not in cmudict</div>";
16}
17print 'Arpabet: <a href="https://en.wikipedia.org/wiki/ARPABET#Symbols">symbol table</a>';
18print "</body></html>\n"
```



Worked example: `CMUDict`

- http://www.speech.cs.cmu.edu/cgi-bin/cmudict

This is how it's served!
Notice it's a cgi-bin!



A worked example: `CMUDict`

- You must have Perl installed
  - **default (macOS/Ubuntu)**: `/usr/bin/perl`
  - **Brew**: `/opt/homebrew/bin/perl`
  - **MacPorts**: `/opt/local/bin/perl`
  - `xcode-select --install` (for macOS only)

CPAN module install:
`/usr/bin/cpan`
`/opt/homebrew/bin/cpan`
`/opt/local/bin/cpan`

- default:
  - `/usr/bin/cpan Lingua::EN::CMUDict`
  - *sometimes also need to* `cpan File::ShareDir`
- if using Brew:
  - install homebrew from http://brew.sh
  - then `brew install perl`

  cpan is the Perl library installer
  - `/opt/homebrew/bin/cpan`
  - `/opt/homebrew/bin/cpan Lingua::EN::CMUDict`



Worked example: `CMUDict`

- Windows 11 PowerShell:
  - `wsl` to start Ubuntu

```
$ which perl
/usr/bin/perl
$ which cpan
/usr/bin/cpan
$ sudo cpan Lingua::EN::CMUDict
$ cd ~
$ mkdir public_html
$ cd public_html
$ cp /mnt/c/Users/username/cmudict.cgi .      (copy from my Windows C drive)
(check permissions for read and execute)
```

Worked example: `CMUDict`

```
#!/usr/bin/perl
use Lingua::EN::CMUDict;
my $obj = new Lingua::EN::CMUDict;
my $string = $ENV{QUERY_STRING};
my $word =  $string ? $string : $ARGV[0];
my $n = $obj->number_of_syllables($word);
print "Content-type: text/html; charset=utf-8\n\n";
print '<html><head><style>div {font-size: x-large}</style></head>';
print '<body><h1>Using CMUDict from Perl</h1><div>';
if ($n) {
  print "<em>$word</em> has $n syllable(s).</div>";
  my $pron = $obj->get_word($word);
  print "$pron<br>";
} else {
  print "<em>$word</em> not in cmudict</div>";
}
print 'Arpabet: <a href="https://en.wikipedia.org/wiki/ARPABET#Symbols">symbol table</a>';
print "</body></html>\n"
```

usually /usr/bin/perl (macOS or Ubuntu) or /opt/homebrew/bin/perl (Homebrew) or /opt/local/bin/perl (MacPorts)

word is in $QUERY_STRING
or on command line $ARGV[0]

recall text/html followed by a *blank line*

Worked example: `CMUDict`

- Command line usage (*assuming I'm in directory* `~/Sites` *or in* `~/public_html`):
- `/usr/bin/perl cmudict.cgi tomato`

```
Content-type: text/html; charset=utf-8

<html><head><style>div {font-size: x-large}</style></head><body><h1>Using
CMUDict from Perl</h1><div><em>tomato</em> has 3 syllable(s).</div>T AH0 - M EY1
- T OW2<br>Arpabet: <a
href="https://en.wikipedia.org/wiki/ARPABET#Symbols">symbol
table</a></body></html>
```



I added `?tomato` in manually
(*cheat for now*)
define a <form> next time!

Worked example: `CMUDict`



Symbols [edit]

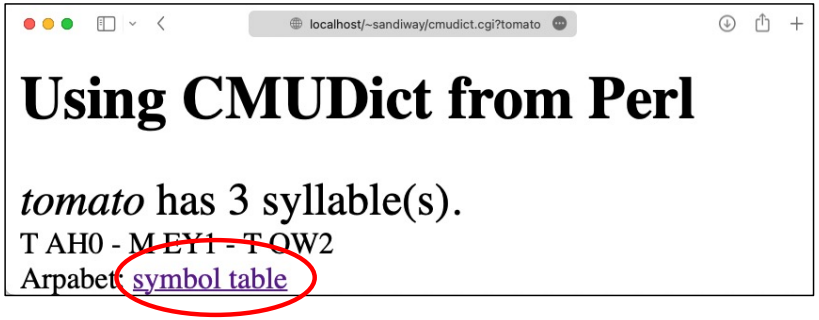
Stress is indicated by a digit immediately following a vowel. Auxiliary symbols are identical in 1- and 2-letter codes. In 2-letter notation, segments are separated by a space.

Vowels[2]

ARPABET 1-letter	ARPABET 2-letter	IPA	Example(s)
a	AA	ɑ~ɒ	balm, bot (with father–bother merger)
@	AE	æ	bat
A	AH	ʌ	butt
c	AO	ɔ	caught, story
W	AW	aʊ	bout
x	AX	ə	comma
–	AXR[3]	ɚ	letter, forward
Y	AY	aɪ	bite
E	EH	ɛ	bet
R	ER	ɝ	bird, foreword
e	EY	eɪ	bait
I	IH	ɪ	bit
X	IX	ɨ	roses, rabbit
i	IY	i	beat
o	OW	oʊ	boat
O	OY	ɔɪ	boy
U	UH	ʊ	book
u	UW	u	boot
–	UX[3]	ʉ	dude

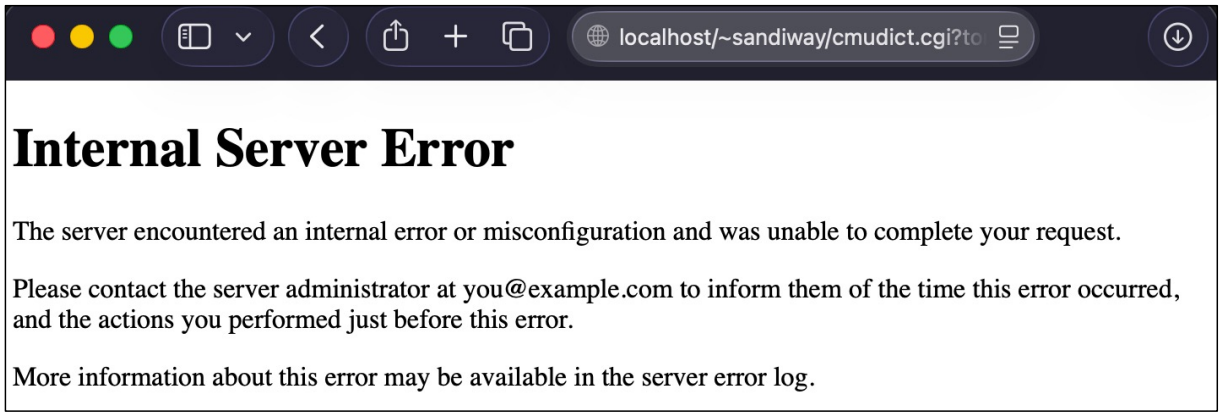
Consonants[2]

ARPABET 1-letter	ARPABET 2-letter	IPA	Example
b	B	b	buy
C	CH	tʃ	China
d	D	d	die
D	DH	ð	thy
F	DX	ɾ	butter
L	EL	l̩	bottle
M	EM	m̩	rhythm
N	EN	n̩	button
f	F	f	fight
g	G	g	guy
h	HH *or* H[3]	h	high
J	JH	dʒ	jive
k	K	k	kite
l	L	l	lie
m	M	m	my
n	N	n	nigh
G	NX *or* NG[3]	ŋ	sing
–	NX[3]	ɾ̃	winner
p	P	p	pie



Worked example: `CMUDict`

- Debugging?
  - look at the log files in the `/var/log/apache2` directory!



Worked example: `CMUDict`

- Debugging? `/opt/homebrew/bin/perl`
  - look at the log files in the `/var/log/apache2` directory!
  - `tail error_log`

```
[Wed Oct 28 19:50:23.926141 2026] [cgi:error] [pid 39364] [client ::1:63912] AH01215: stderr from
/Users/sandiway/Sites/cmudict.cgi: Can't locate Lingua/EN/CMUDict.pm in @INC (you may need to install
the Lingua::EN::CMUDict module) (@INC entries checked:
/opt/homebrew/opt/perl/lib/perl5/site_perl/5.44/darwin-thread-multi-2level
/opt/homebrew/opt/perl/lib/perl5/site_perl/5.44 /opt/homebrew/lib/perl5/vendor_perl/5.44/darwin-thread-
multi-2level /opt/homebrew/lib/perl5/vendor_perl/5.44 /opt/homebrew/opt/perl/lib/perl5/5.44/darwin-
thread-multi-2level /opt/homebrew/opt/perl/lib/perl5/5.44 /opt/homebrew/lib/perl5/site_perl/5.44) at
/Users/sandiway/Sites/cmudict.cgi line 2.

[Wed Oct 28 19:50:23.927411 2026] [cgi:error] [pid 39364] [client ::1:63912] AH01215: stderr from
/Users/sandiway/Sites/cmudict.cgi: BEGIN failed--compilation aborted at
/Users/sandiway/Sites/cmudict.cgi line 2.

[Wed Oct 28 19:50:23.927484 2026] [cgi:error] [pid 39364] [client ::1:63912] End of script output
before headers: cmudict.cgi
```

@INC is different! I have the CMUDict.pm in ~/perl5/lib/perl5/Lingua/EN

Worked example: `CMUDict`

On recent macOSs
`/System/Library` locked out for installs
`/Library/Perl` is left open

- Debugging? `/usr/bin/perl`
  - look at the log files in the `/var/log/apache2` directory!
  - `tail error_log`

```
[Wed Oct 28 20:00:21.111092 2026] [cgi:error] [pid 39364] [client ::1:64043] AH01215: stderr from
/Users/sandiway/Sites/cmudict.cgi: Can't locate Lingua/EN/CMUDict.pm in @INC (you may need to install
the Lingua::EN::CMUDict module) (@INC contains: /Library/Perl/5.34/darwin-thread-multi-2level
/Library/Perl/5.34 /Network/Library/Perl/5.34/darwin-thread-multi-2level /Network/Library/Perl/5.34
/Library/Perl/Updates/5.34.1 /System/Library/Perl/5.34/darwin-thread-multi-2level
/System/Library/Perl/5.34 /System/Library/Perl/Extras/5.34/darwin-thread-multi-2level
/System/Library/Perl/Extras/5.34) at /Users/sandiway/Sites/cmudict.cgi line 2.

[Wed Oct 28 20:00:21.111644 2026] [cgi:error] [pid 39364] [client ::1:64043] AH01215: stderr from
/Users/sandiway/Sites/cmudict.cgi: BEGIN failed--compilation aborted at
/Users/sandiway/Sites/cmudict.cgi line 2.

[Wed Oct 28 20:00:21.111677 2026] [cgi:error] [pid 39364] [client ::1:64043] End of script output
before headers: cmudict.cgi
```

@INC is different again!

Worked example: `CMUDict`

- **sudo** `/usr/bin/cpan Lingua::EN::CMUDict` will target `/Library/Perl`!

```
$ sudo /usr/bin/cpan Lingua::EN::CMUDict
Password:
Loading internal logger. Log::Log4perl recommended for better logging
Reading '/Users/sandiway/.cpan/Metadata'
  Database was generated on Thu, 29 Oct 2026 00:54:04 GMT
Running install for module 'Lingua::EN::CMUDict'
…
Installing /Library/Perl/5.34/auto/share/dist/Lingua-EN-CMUDict/cmusyldict.db
Installing /Library/Perl/5.34/Lingua/EN/CMUDict.pm
Installing /usr/local/share/man/man3/Lingua::EN::CMUDict.3pm
Appending installation info to /Library/Perl/Updates/5.34.1/darwin-thread-multi-
2level/perllocal.pod
  LMETCALF/Lingua-EN-CMUDict-0.07.tar.gz
  /usr/bin/make install  -- OK
```

Worked example: CMUDict

- `tail /var/log/apache2/access_log`
- `::1 - - [28/Oct/2026:20:21:39 -0700] "GET /~sandiway/test.cgi HTTP/1.1" 200 1057`
- `::1 - - [28/Oct/2026:20:21:56 -0700] "GET /~sandiway/cmudict.cgi?tomato HTTP/1.1" 200 233`

Worked example: `CMUDict`

https://metacpan.org/pod/Lingua::EN::CMUDict

Lingua::EN::CMUDict - Perl extension for utilizing the CMU dictionary file

SYNOPSIS

```
use Lingua::EN::CMUDict;
my $obj = new Lingua::EN::CMUDict;
print $obj->number_of_syllables("test");
```

DESCRIPTION

This version of the CMU Pronouncing dictionary was generated from the original dictionary and designed to syllabify it. The paper *On the Syllabification of Phonemes* by Susan Bartlett, Grzegorz Kondrak and Colin Cherry (NAACL-HLT 2009) covers the methods used to generate the dictionary.

EXPORT

None by default.

METHODS

new(cmudict=>*file*)

Creates a new object, populating it with the cmusyldict db file. If the cmudict argument is passed with a filename as the argument, that file is used. If you do not use that argument, the default cmusyldict db file installed with the module is used.

rhymes(word)

In the case of an array being returned, returns all rhymes to the given word. In a scalar context, returns a single rhyme.

number_of_syllables(word)

Returns the number of syllables in the word. Many pluralities do not add syllable counts and are therefore not in the original database. This code tries to be intelligent by looking for those and returning the number of syllables. Also, if a sentence is passed in, returns the number of syllables in the sentence. Doesn't currently deal with punctuation very well.

get_word(word)

Returns the pronunciation for the word with syllable boundaries.

SEE ALSO

Lingua::EN::Phoneme -- another way of accessing the CMU Pronunciation dictionary.

AUTHOR

Leigh Metcalf, <leigh@fprime.net>

COPYRIGHT AND LICENSE

Copyright (C) 2011 by Leigh Metcalf

Worked example: `CMUDict`

- Actual entry in CMUDict looks like this:

```
1. TOMATO T AH0 M EY1 T OW2
2. TOMATO(1) T AH0 M AA1 T OW2
3. TOMATOE T AH0 M EY1 T OW0
4. TOMATOE(1) T AH0 M AA1 T OW0
5. TOMATOES T AH0 M EY1 T OW0 Z
6. TOMATOES(1) T AH0 M AA1 T OW0 Z
7. TOMATOS T AH0 M EY1 T OW2 Z
8. TOMATOS(1) T AH0 M AA1 T OW2 Z
```



A shame the code doesn't retrieve all available pronounciations!

Worked example: `CMUDict`

```
~$ python3
Python 3.14.7 (v3.14.7:823f0323ee6, Aug  5 2026, 07:07:01) [Clang 21.0.0 (clang-
2100.1.1.101)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> from nltk.corpus import cmudict
>>> d = cmudict.dict()
>>> d['tomato']
[['T', 'AH0', 'M', 'EY1', 'T', 'OW2'], ['T', 'AH0', 'M', 'AA1', 'T', 'OW2']]
>>> d['trump']
[['T', 'R', 'AH1', 'M', 'P']]
>>> d['obama']
[['OW2', 'B', 'AA1', 'M', 'AH0']]
>>> [[ph for ph in p if ph[-1].isdigit()] for p in d['obama']]
[['OW2', 'AA1', 'AH0']]
>>> [[ph for ph in p if ph[-1].isdigit()] for p in d['tomato']]
[['AH0', 'EY1', 'OW2'], ['AH0', 'AA1', 'OW2']]
>>> [len([ph for ph in p if ph[-1].isdigit()]) for p in d['tomato']]
[3, 3]
>>>
```