GNU Wget

The noninteractive downloading utility
Updated for Wget 1.8.1, December 2001

by Hrvoje Nikšić and the developers

1 Overview

GNU Wget is a free utility for non-interactive download of files from the Web. It supports HTTP, HTTPS, and FTP protocols, as well as retrieval through HTTP proxies.

This chapter is a partial overview of Wget's features.

- Wget is non-interactive, meaning that it can work in the background, while the user is not logged on. This allows you to start a retrieval and disconnect from the system, letting Wget finish the work. By contrast, most of the Web browsers require constant user's presence, which can be a great hindrance when transferring a lot of data.
- Wget can follow links in HTML pages and create local versions of remote web sites, fully recreating the directory structure of the original site. This is sometimes referred to as "recursive downloading." While doing that, Wget respects the Robot Exclusion Standard (`/robots.txt`). Wget can be instructed to convert the links in downloaded HTML files to the local files for offline viewing.
- File name wildcard matching and recursive mirroring of directories are available when retrieving via FTP. Wget can read the time-stamp information given by both HTTP and FTP servers, and store it locally. Thus Wget can see if the remote file has changed since last retrieval, and automatically retrieve the new version if it has. This makes Wget suitable for mirroring of FTP sites, as well as home pages.
- Wget has been designed for robustness over slow or unstable network connections; if a download fails due to a network problem, it will keep retrying until the whole file has been retrieved. If the server supports regetting, it will instruct the server to continue the download from where it left off.
- Wget supports proxy servers, which can lighten the network load, speed up retrieval and provide access behind firewalls. However, if you are behind a firewall that requires that you use a socks style gateway, you can get the socks library and build Wget with support for socks. Wget also supports the passive FTP downloading as an option.
- Builtin features offer mechanisms to tune which links you wish to follow (see Chapter 4 [Following Links], page 18).
- The retrieval is conveniently traced with printing dots, each dot representing a fixed amount of data received (1KB by default). These representations can be customized to your preferences.
- Most of the features are fully configurable, either through command line options, or via the initialization file `.wgetrc` (see Chapter 6 [Startup File], page 25). Wget allows you to define *global* startup files (`/usr/local/etc/wgetrc` by default) for site settings.
- Finally, GNU Wget is free software. This means that everyone may use it, redistribute it and/or modify it under the terms of the GNU General Public License, as published by the Free Software Foundation (see Chapter 10 [Copying], page 42).

2 Invoking

By default, Wget is very simple to invoke. The basic syntax is:

```
wget [option]... [URL]...
```

Wget will simply download all the URLs specified on the command line. *URL* is a *Uniform Resource Locator*, as defined below.

However, you may wish to change some of the default parameters of Wget. You can do it two ways: permanently, adding the appropriate command to '.wgetrc' (see Chapter 6 [Startup File], page 25), or specifying it on the command line.

2.1 URL Format

URL is an acronym for Uniform Resource Locator. A uniform resource locator is a compact string representation for a resource available via the Internet. Wget recognizes the URL syntax as per RFC1738. This is the most widely used form (square brackets denote optional parts):

```
http://host[:port]/directory/file
ftp://host[:port]/directory/file
```

You can also encode your username and password within a URL:

```
ftp://user:password@host/path
http://user:password@host/path
```

Either *user* or *password*, or both, may be left out. If you leave out either the HTTP username or password, no authentication will be sent. If you leave out the FTP username, '`anonymous`' will be used. If you leave out the FTP password, your email address will be supplied as a default password.[1]

You can encode unsafe characters in a URL as '`%xy`', `xy` being the hexadecimal representation of the character's ASCII value. Some common unsafe characters include '`%`' (quoted as '`%25`'), '`:`' (quoted as '`%3A`'), and '`@`' (quoted as '`%40`'). Refer to RFC1738 for a comprehensive list of unsafe characters.

Wget also supports the `type` feature for FTP URLs. By default, FTP documents are retrieved in the binary mode (type '`i`'), which means that they are downloaded unchanged. Another useful mode is the '`a`' (*ASCII*) mode, which converts the line delimiters between the different operating systems, and is thus useful for text files. Here is an example:

```
ftp://host/directory/file;type=a
```

Two alternative variants of URL specification are also supported, because of historical (hysterical?) reasons and their widespreaded use.

[1] If you have a '`.netrc`' file in your home directory, password will also be searched for there.

FTP-only syntax (supported by NcFTP):

```
host:/dir/file
```

HTTP-only syntax (introduced by Netscape):

```
host[:port]/dir/file
```

These two alternative forms are deprecated, and may cease being supported in the future.

If you do not understand the difference between these notations, or do not know which one to use, just use the plain ordinary format you use with your favorite browser, like Lynx or Netscape.

2.2 Option Syntax

Since Wget uses GNU getopts to process its arguments, every option has a short form and a long form. Long options are more convenient to remember, but take time to type. You may freely mix different option styles, or specify options after the command-line arguments. Thus you may write:

```
wget -r --tries=10 http://fly.srk.fer.hr/ -o log
```

The space between the option accepting an argument and the argument may be omitted. Instead '-o log' you can write '-olog'.

You may put several options that do not require arguments together, like:

```
wget -drc URL
```

This is a complete equivalent of:

```
wget -d -r -c URL
```

Since the options can be specified after the arguments, you may terminate them with '--'. So the following will try to download URL '-x', reporting failure to 'log':

```
wget -o log -- -x
```

The options that accept comma-separated lists all respect the convention that specifying an empty list clears its value. This can be useful to clear the '.wgetrc' settings. For instance, if your '.wgetrc' sets exclude_directories to '/cgi-bin', the following example will first reset it, and then set it to exclude '/~nobody' and '/~somebody'. You can also clear the lists in '.wgetrc' (see Section 6.2 [Wgetrc Syntax], page 25).

```
wget -X '' -X /~nobody,/~somebody
```

2.3 Basic Startup Options

'-V'
'--version'
Display the version of Wget.

'-h'
'--help' Print a help message describing all of Wget's command-line options.

'-b'
'--background'
Go to background immediately after startup. If no output file is specified via the '-o', output is redirected to 'wget-log'.

'-e *command*'
'--execute *command*'
Execute *command* as if it were a part of '.wgetrc' (see Chapter 6 [Startup File], page 25). A command thus invoked will be executed *after* the commands in '.wgetrc', thus taking precedence over them.

2.4 Logging and Input File Options

'-o *logfile*'
'--output-file=*logfile*'
Log all messages to *logfile*. The messages are normally reported to standard error.

'-a *logfile*'
'--append-output=*logfile*'
Append to *logfile*. This is the same as '-o', only it appends to *logfile* instead of overwriting the old log file. If *logfile* does not exist, a new file is created.

'-d'
'--debug' Turn on debug output, meaning various information important to the developers of Wget if it does not work properly. Your system administrator may have chosen to compile Wget without debug support, in which case '-d' will not work. Please note that compiling with debug support is always safe—Wget compiled with the debug support will *not* print any debug info unless requested with '-d'. See Section 8.4 [Reporting Bugs], page 37, for more information on how to use '-d' for sending bug reports.

'-q'
'--quiet' Turn off Wget's output.

'-v'
'--verbose'
Turn on verbose output, with all the available data. The default output is verbose.

'-nv'
'--non-verbose'
Non-verbose output—turn off verbose without being completely quiet (use '-q' for that), which means that error messages and basic information still get printed.

'-i *file*'
'--input-file=*file*'
Read URLs from *file*, in which case no URLs need to be on the command line. If there are URLs both on the command line and in an input file, those on the command lines

will be the first ones to be retrieved. The *file* need not be an HTML document (but no harm if it is)—it is enough if the URLs are just listed sequentially.

However, if you specify '`--force-html`', the document will be regarded as '`html`'. In that case you may have problems with relative links, which you can solve either by adding `<base href="`*url*`">` to the documents or by specifying '`--base=`*url*' on the command line.

'`-F`'
'`--force-html`'
: When input is read from a file, force it to be treated as an HTML file. This enables you to retrieve relative links from existing HTML files on your local disk, by adding `<base href="`*url*`">` to HTML, or using the '`--base`' command-line option.

'`-B` *URL*'
'`--base=`*URL*'
: When used in conjunction with '`-F`', prepends *URL* to relative links in the file specified by '`-i`'.

2.5 Download Options

'`--bind-address=`*ADDRESS*'
: When making client TCP/IP connections, `bind()` to *ADDRESS* on the local machine. *ADDRESS* may be specified as a hostname or IP address. This option can be useful if your machine is bound to multiple IPs.

'`-t` *number*'
'`--tries=`*number*'
: Set number of retries to *number*. Specify 0 or '`inf`' for infinite retrying.

'`-O` *file*'
'`--output-document=`*file*'
: The documents will not be written to the appropriate files, but all will be concatenated together and written to *file*. If *file* already exists, it will be overwritten. If the *file* is '`-`', the documents will be written to standard output. Including this option automatically sets the number of tries to 1.

'`-nc`'
'`--no-clobber`'
: If a file is downloaded more than once in the same directory, Wget's behavior depends on a few options, including '`-nc`'. In certain cases, the local file will be *clobbered*, or overwritten, upon repeated download. In other cases it will be preserved.

When running Wget without '`-N`', '`-nc`', or '`-r`', downloading the same file in the same directory will result in the original copy of *file* being preserved and the second copy being named '*file*`.1`'. If that file is downloaded yet again, the third copy will be named '*file*`.2`', and so on. When '`-nc`' is specified, this behavior is suppressed, and Wget will refuse to download newer copies of '*file*'. Therefore, "`no-clobber`" is actually a misnomer in this mode—it's not clobbering that's prevented (as the numeric suffixes were already preventing clobbering), but rather the multiple version saving that's prevented.

When running Wget with '`-r`', but without '`-N`' or '`-nc`', re-downloading a file will result in the new copy simply overwriting the old. Adding '`-nc`' will prevent this behavior, instead causing the original version to be preserved and any newer copies on the server to be ignored.

When running Wget with '-N', with or without '-r', the decision as to whether or not to download a newer copy of a file depends on the local and remote timestamp and size of the file (see Chapter 5 [Time-Stamping], page 22). '-nc' may not be specified at the same time as '-N'.

Note that when '-nc' is specified, files with the suffixes '.html' or (yuck) '.htm' will be loaded from the local disk and parsed as if they had been retrieved from the Web.

'-c'
'--continue'

Continue getting a partially-downloaded file. This is useful when you want to finish up a download started by a previous instance of Wget, or by another program. For instance:

```
wget -c ftp://sunsite.doc.ic.ac.uk/ls-lR.Z
```

If there is a file named 'ls-lR.Z' in the current directory, Wget will assume that it is the first portion of the remote file, and will ask the server to continue the retrieval from an offset equal to the length of the local file.

Note that you don't need to specify this option if you just want the current invocation of Wget to retry downloading a file should the connection be lost midway through. This is the default behavior. '-c' only affects resumption of downloads started *prior* to this invocation of Wget, and whose local files are still sitting around.

Without '-c', the previous example would just download the remote file to 'ls-lR.Z.1', leaving the truncated 'ls-lR.Z' file alone.

Beginning with Wget 1.7, if you use '-c' on a non-empty file, and it turns out that the server does not support continued downloading, Wget will refuse to start the download from scratch, which would effectively ruin existing contents. If you really want the download to start from scratch, remove the file.

Also beginning with Wget 1.7, if you use '-c' on a file which is of equal size as the one on the server, Wget will refuse to download the file and print an explanatory message. The same happens when the file is smaller on the server than locally (presumably because it was changed on the server since your last download attempt)—because "continuing" is not meaningful, no download occurs.

On the other side of the coin, while using '-c', any file that's bigger on the server than locally will be considered an incomplete download and only (length(remote) - length(local)) bytes will be downloaded and tacked onto the end of the local file. This behavior can be desirable in certain cases—for instance, you can use 'wget -c' to download just the new portion that's been appended to a data collection or log file.

However, if the file is bigger on the server because it's been *changed*, as opposed to just *appended* to, you'll end up with a garbled file. Wget has no way of verifying that the local file is really a valid prefix of the remote file. You need to be especially careful of this when using '-c' in conjunction with '-r', since every file will be considered as an "incomplete download" candidate.

Another instance where you'll get a garbled file if you try to use '-c' is if you have a lame HTTP proxy that inserts a "transfer interrupted" string into the local file. In the future a "rollback" option may be added to deal with this case.

Note that '-c' only works with FTP servers and with HTTP servers that support the Range header.

'--progress=*type*'

Select the type of the progress indicator you wish to use. Legal indicators are "dot" and "bar".

The "dot" indicator is used by default. It traces the retrieval by printing dots on the screen, each dot representing a fixed amount of downloaded data.

When using the dotted retrieval, you may also set the *style* by specifying the type as '`dot:`*style*'. Different styles assign different meaning to one dot. With the `default` style each dot represents 1K, there are ten dots in a cluster and 50 dots in a line. The `binary` style has a more "computer"-like orientation—8K dots, 16-dots clusters and 48 dots per line (which makes for 384K lines). The `mega` style is suitable for downloading very large files—each dot represents 64K retrieved, there are eight dots in a cluster, and 48 dots on each line (so each line contains 3M).

Specifying '`--progress=bar`' will draw a nice ASCII progress bar graphics (a.k.a "thermometer" display) to indicate retrieval. If the output is not a TTY, this option will be ignored, and Wget will revert to the dot indicator. If you want to force the bar indicator, use '`--progress=bar:force`'.

'`-N`'
'`--timestamping`'
: Turn on time-stamping. See Chapter 5 [Time-Stamping], page 22, for details.

'`-S`'
'`--server-response`'
: Print the headers sent by HTTP servers and responses sent by FTP servers.

'`--spider`'
: When invoked with this option, Wget will behave as a Web *spider*, which means that it will not download the pages, just check that they are there. You can use it to check your bookmarks, e.g. with:

```
wget --spider --force-html -i bookmarks.html
```

This feature needs much more work for Wget to get close to the functionality of real WWW spiders.

'`-T seconds`'
'`--timeout=`*seconds*'
: Set the read timeout to *seconds* seconds. Whenever a network read is issued, the file descriptor is checked for a timeout, which could otherwise leave a pending connection (uninterrupted read). The default timeout is 900 seconds (fifteen minutes). Setting timeout to 0 will disable checking for timeouts.

Please do not lower the default timeout value with this option unless you know what you are doing.

'`-w` *seconds*'
'`--wait=`*seconds*'
: Wait the specified number of seconds between the retrievals. Use of this option is recommended, as it lightens the server load by making the requests less frequent. Instead of in seconds, the time can be specified in minutes using the `m` suffix, in hours using `h` suffix, or in days using `d` suffix.

Specifying a large value for this option is useful if the network or the destination host is down, so that Wget can wait long enough to reasonably expect the network error to be fixed before the retry.

'`--waitretry=`*seconds*'
: If you don't want Wget to wait between *every* retrieval, but only between retries of failed downloads, you can use this option. Wget will use *linear backoff*, waiting 1 second after the first failure on a given file, then waiting 2 seconds after the second failure on that file, up to the maximum number of *seconds* you specify. Therefore, a value of 10 will actually make Wget wait up to (1 + 2 + ... + 10) = 55 seconds per file.

Note that this option is turned on by default in the global '`wgetrc`' file.

'--random-wait'
: Some web sites may perform log analysis to identify retrieval programs such as Wget by looking for statistically significant similarities in the time between requests. This option causes the time between requests to vary between 0 and 2 * *wait* seconds, where *wait* was specified using the '-w' or '--wait' options, in order to mask Wget's presence from such analysis.

  A recent article in a publication devoted to development on a popular consumer platform provided code to perform this analysis on the fly. Its author suggested blocking at the class C address level to ensure automated retrieval programs were blocked despite changing DHCP-supplied addresses.

  The '--random-wait' option was inspired by this ill-advised recommendation to block many unrelated users from a web site due to the actions of one.

'-Y on/off'
'--proxy=on/off'
: Turn proxy support on or off. The proxy is on by default if the appropriate environmental variable is defined.

'-Q *quota*'
'--quota=*quota*'
: Specify download quota for automatic retrievals. The value can be specified in bytes (default), kilobytes (with 'k' suffix), or megabytes (with 'm' suffix).

  Note that quota will never affect downloading a single file. So if you specify 'wget -Q10k ftp://wuarchive.wustl.edu/ls-lR.gz', all of the 'ls-lR.gz' will be downloaded. The same goes even when several URLs are specified on the command-line. However, quota is respected when retrieving either recursively, or from an input file. Thus you may safely type 'wget -Q2m -i sites'—download will be aborted when the quota is exceeded.

  Setting quota to 0 or to 'inf' unlimits the download quota.

2.6 Directory Options

'-nd'
'--no-directories'
: Do not create a hierarchy of directories when retrieving recursively. With this option turned on, all files will get saved to the current directory, without clobbering (if a name shows up more than once, the filenames will get extensions '.n').

'-x'
'--force-directories'
: The opposite of '-nd'—create a hierarchy of directories, even if one would not have been created otherwise. E.g. 'wget -x http://fly.srk.fer.hr/robots.txt' will save the downloaded file to 'fly.srk.fer.hr/robots.txt'.

'-nH'
'--no-host-directories'
: Disable generation of host-prefixed directories. By default, invoking Wget with '-r http://fly.srk.fer.hr/' will create a structure of directories beginning with 'fly.srk.fer.hr/'. This option disables such behavior.

'--cut-dirs=*number*'
: Ignore *number* directory components. This is useful for getting a fine-grained control over the directory where recursive retrieval will be saved.

Take, for example, the directory at '`ftp://ftp.xemacs.org/pub/xemacs/`'. If you retrieve it with '`-r`', it will be saved locally under '`ftp.xemacs.org/pub/xemacs/`'. While the '`-nH`' option can remove the '`ftp.xemacs.org/`' part, you are still stuck with '`pub/xemacs`'. This is where '`--cut-dirs`' comes in handy; it makes Wget not "see" *number* remote directory components. Here are several examples of how '`--cut-dirs`' option works.

```
No options        -> ftp.xemacs.org/pub/xemacs/
-nH               -> pub/xemacs/
-nH --cut-dirs=1  -> xemacs/
-nH --cut-dirs=2  -> .

--cut-dirs=1      -> ftp.xemacs.org/xemacs/
...
```

If you just want to get rid of the directory structure, this option is similar to a combination of '`-nd`' and '`-P`'. However, unlike '`-nd`', '`--cut-dirs`' does not lose with subdirectories—for instance, with '`-nH --cut-dirs=1`', a '`beta/`' subdirectory will be placed to '`xemacs/beta`', as one would expect.

'`-P` *prefix*'
'`--directory-prefix=`*prefix*'

Set directory prefix to *prefix*. The *directory prefix* is the directory where all other files and subdirectories will be saved to, i.e. the top of the retrieval tree. The default is '`.`' (the current directory).

2.7 HTTP Options

'`-E`'
'`--html-extension`'

If a file of type '`text/html`' is downloaded and the URL does not end with the regexp '`\.[Hh][Tt][Mm][Ll]?`', this option will cause the suffix '`.html`' to be appended to the local filename. This is useful, for instance, when you're mirroring a remote site that uses '`.asp`' pages, but you want the mirrored pages to be viewable on your stock Apache server. Another good use for this is when you're downloading the output of CGIs. A URL like '`http://site.com/article.cgi?25`' will be saved as '`article.cgi?25.html`'.

Note that filenames changed in this way will be re-downloaded every time you re-mirror a site, because Wget can't tell that the local '*X*`.html`' file corresponds to remote URL '*X*' (since it doesn't yet know that the URL produces output of type '`text/html`'. To prevent this re-downloading, you must use '`-k`' and '`-K`' so that the original version of the file will be saved as '*X*`.orig`' (see Section 2.9 [Recursive Retrieval Options], page 12).

'`--http-user=`*user*'
'`--http-passwd=`*password*'

Specify the username *user* and password *password* on an HTTP server. According to the type of the challenge, Wget will encode them using either the `basic` (insecure) or the `digest` authentication scheme.

Another way to specify username and password is in the URL itself (see Section 2.1 [URL Format], page 2). For more information about security issues with Wget, See Section 9.2 [Security Considerations], page 40.

'-C on/off'
'--cache=on/off'
: When set to off, disable server-side cache. In this case, Wget will send the remote server an appropriate directive ('Pragma: no-cache') to get the file from the remote service, rather than returning the cached version. This is especially useful for retrieving and flushing out-of-date documents on proxy servers.

  Caching is allowed by default.

'--cookies=on/off'
: When set to off, disable the use of cookies. Cookies are a mechanism for maintaining server-side state. The server sends the client a cookie using the Set-Cookie header, and the client responds with the same cookie upon further requests. Since cookies allow the server owners to keep track of visitors and for sites to exchange this information, some consider them a breach of privacy. The default is to use cookies; however, *storing* cookies is not on by default.

'--load-cookies *file*'
: Load cookies from *file* before the first HTTP retrieval. *file* is a textual file in the format originally used by Netscape's 'cookies.txt' file.

  You will typically use this option when mirroring sites that require that you be logged in to access some or all of their content. The login process typically works by the web server issuing an HTTP cookie upon receiving and verifying your credentials. The cookie is then resent by the browser when accessing that part of the site, and so proves your identity.

  Mirroring such a site requires Wget to send the same cookies your browser sends when communicating with the site. This is achieved by '--load-cookies'—simply point Wget to the location of the 'cookies.txt' file, and it will send the same cookies your browser would send in the same situation. Different browsers keep textual cookie files in different locations:

  Netscape 4.x.
  : The cookies are in '~/.netscape/cookies.txt'.

  Mozilla and Netscape 6.x.
  : Mozilla's cookie file is also named 'cookies.txt', located somewhere under '~/.mozilla', in the directory of your profile. The full path usually ends up looking somewhat like '~/.mozilla/default/*some-weird-string*/cookies.txt'.

  Internet Explorer.
  : You can produce a cookie file Wget can use by using the File menu, Import and Export, Export Cookies. This has been tested with Internet Explorer 5; it is not guaranteed to work with earlier versions.

  Other browsers.
  : If you are using a different browser to create your cookies, '--load-cookies' will only work if you can locate or produce a cookie file in the Netscape format that Wget expects.

  If you cannot use '--load-cookies', there might still be an alternative. If your browser supports a "cookie manager", you can use it to view the cookies used when accessing the site you're mirroring. Write down the name and value of the cookie, and manually instruct Wget to send those cookies, bypassing the "official" cookie support:

```
wget --cookies=off --header "Cookie: name=value"
```

'--save-cookies *file*'
: Save cookies from *file* at the end of session. Cookies whose expiry time is not specified, or those that have already expired, are not saved.

'--ignore-length'
: Unfortunately, some HTTP servers (CGI programs, to be more precise) send out bogus Content-Length headers, which makes Wget go wild, as it thinks not all the document was retrieved. You can spot this syndrome if Wget retries getting the same document again and again, each time claiming that the (otherwise normal) connection has closed on the very same byte.

  With this option, Wget will ignore the Content-Length header—as if it never existed.

'--header=*additional-header*'
: Define an *additional-header* to be passed to the HTTP servers. Headers must contain a ':' preceded by one or more non-blank characters, and must not contain newlines.

  You may define more than one additional header by specifying '--header' more than once.

```
wget --header='Accept-Charset: iso-8859-2' \
     --header='Accept-Language: hr'        \
       http://fly.srk.fer.hr/
```

  Specification of an empty string as the header value will clear all previous user-defined headers.

'--proxy-user=*user*'
'--proxy-passwd=*password*'
: Specify the username *user* and password *password* for authentication on a proxy server. Wget will encode them using the basic authentication scheme.

'--referer=*url*'
: Include 'Referer: *url*' header in HTTP request. Useful for retrieving documents with server-side processing that assume they are always being retrieved by interactive web browsers and only come out properly when Referer is set to one of the pages that point to them.

'-s'
'--save-headers'
: Save the headers sent by the HTTP server to the file, preceding the actual contents, with an empty line as the separator.

'-U *agent-string*'
'--user-agent=*agent-string*'
: Identify as *agent-string* to the HTTP server.

  The HTTP protocol allows the clients to identify themselves using a User-Agent header field. This enables distinguishing the WWW software, usually for statistical purposes or for tracing of protocol violations. Wget normally identifies as 'Wget/*version*', *version* being the current version number of Wget.

  However, some sites have been known to impose the policy of tailoring the output according to the User-Agent-supplied information. While conceptually this is not such a bad idea, it has been abused by servers denying information to clients other than Mozilla or Microsoft Internet Explorer. This option allows you to change the User-Agent line issued by Wget. Use of this option is discouraged, unless you really know what you are doing.

2.8 FTP Options

'-nr'
'--dont-remove-listing'

Don't remove the temporary '.listing' files generated by FTP retrievals. Normally, these files contain the raw directory listings received from FTP servers. Not removing them can be useful for debugging purposes, or when you want to be able to easily check on the contents of remote server directories (e.g. to verify that a mirror you're running is complete).

Note that even though Wget writes to a known filename for this file, this is not a security hole in the scenario of a user making '.listing' a symbolic link to '/etc/passwd' or something and asking root to run Wget in his or her directory. Depending on the options used, either Wget will refuse to write to '.listing', making the globbing/recursion/time-stamping operation fail, or the symbolic link will be deleted and replaced with the actual '.listing' file, or the listing will be written to a '.listing.*number*' file.

Even though this situation isn't a problem, though, root should never run Wget in a non-trusted user's directory. A user could do something as simple as linking 'index.html' to '/etc/passwd' and asking root to run Wget with '-N' or '-r' so the file will be overwritten.

'-g on/off'
'--glob=on/off'

Turn FTP globbing on or off. Globbing means you may use the shell-like special characters (*wildcards*), like '*', '?', '[' and ']' to retrieve more than one file from the same directory at once, like:

```
wget ftp://gnjilux.srk.fer.hr/*.msg
```

By default, globbing will be turned on if the URL contains a globbing character. This option may be used to turn globbing on or off permanently.

You may have to quote the URL to protect it from being expanded by your shell. Globbing makes Wget look for a directory listing, which is system-specific. This is why it currently works only with Unix FTP servers (and the ones emulating Unix ls output).

'--passive-ftp'

Use the *passive* FTP retrieval scheme, in which the client initiates the data connection. This is sometimes required for FTP to work behind firewalls.

'--retr-symlinks'

Usually, when retrieving FTP directories recursively and a symbolic link is encountered, the linked-to file is not downloaded. Instead, a matching symbolic link is created on the local filesystem. The pointed-to file will not be downloaded unless this recursive retrieval would have encountered it separately and downloaded it anyway.

When '--retr-symlinks' is specified, however, symbolic links are traversed and the pointed-to files are retrieved. At this time, this option does not cause Wget to traverse symlinks to directories and recurse through them, but in the future it should be enhanced to do this.

Note that when retrieving a file (not a directory) because it was specified on the commandline, rather than because it was recursed to, this option has no effect. Symbolic links are always traversed in this case.

2.9 Recursive Retrieval Options

'-r'
'--recursive'
: Turn on recursive retrieving. See Chapter 3 [Recursive Retrieval], page 17, for more details.

'-l *depth*'
'--level=*depth*'
: Specify recursion maximum depth level *depth* (see Chapter 3 [Recursive Retrieval], page 17). The default maximum depth is 5.

'--delete-after'
: This option tells Wget to delete every single file it downloads, *after* having done so. It is useful for pre-fetching popular pages through a proxy, e.g.:

```
wget -r -nd --delete-after http://whatever.com/~popular/page/
```

The '-r' option is to retrieve recursively, and '-nd' to not create directories.

Note that '--delete-after' deletes files on the local machine. It does not issue the 'DELE' command to remote FTP sites, for instance. Also note that when '--delete-after' is specified, '--convert-links' is ignored, so '.orig' files are simply not created in the first place.

'-k'
'--convert-links'
: After the download is complete, convert the links in the document to make them suitable for local viewing. This affects not only the visible hyperlinks, but any part of the document that links to external content, such as embedded images, links to style sheets, hyperlinks to non-HTML content, etc.

Each link will be changed in one of the two ways:

- The links to files that have been downloaded by Wget will be changed to refer to the file they point to as a relative link.

  Example: if the downloaded file '/foo/doc.html' links to '/bar/img.gif', also downloaded, then the link in 'doc.html' will be modified to point to '../bar/img.gif'. This kind of transformation works reliably for arbitrary combinations of directories.

- The links to files that have not been downloaded by Wget will be changed to include host name and absolute path of the location they point to.

  Example: if the downloaded file '/foo/doc.html' links to '/bar/img.gif' (or to '../bar/img.gif'), then the link in 'doc.html' will be modified to point to 'http://*hostname*/bar/img.gif'.

Because of this, local browsing works reliably: if a linked file was downloaded, the link will refer to its local name; if it was not downloaded, the link will refer to its full Internet address rather than presenting a broken link. The fact that the former links are converted to relative links ensures that you can move the downloaded hierarchy to another directory.

Note that only at the end of the download can Wget know which links have been downloaded. Because of that, the work done by '-k' will be performed at the end of all the downloads.

'-K'
'--backup-converted'
: When converting a file, back up the original version with a '.orig' suffix. Affects the behavior of '-N' (see Section 5.2 [HTTP Time-Stamping Internals], page 23).

'-m'
'--mirror'

Turn on options suitable for mirroring. This option turns on recursion and time-stamping, sets infinite recursion depth and keeps FTP directory listings. It is currently equivalent to '-r -N -l inf -nr'.

'-p'
'--page-requisites'

This option causes Wget to download all the files that are necessary to properly display a given HTML page. This includes such things as inlined images, sounds, and referenced stylesheets.

Ordinarily, when downloading a single HTML page, any requisite documents that may be needed to display it properly are not downloaded. Using '-r' together with '-l' can help, but since Wget does not ordinarily distinguish between external and inlined documents, one is generally left with "leaf documents" that are missing their requisites.

For instance, say document '1.html' contains an <IMG> tag referencing '1.gif' and an <A> tag pointing to external document '2.html'. Say that '2.html' is similar but that its image is '2.gif' and it links to '3.html'. Say this continues up to some arbitrarily high number.

If one executes the command:

```
wget -r -l 2 http://site/1.html
```

then '1.html', '1.gif', '2.html', '2.gif', and '3.html' will be downloaded. As you can see, '3.html' is without its requisite '3.gif' because Wget is simply counting the number of hops (up to 2) away from '1.html' in order to determine where to stop the recursion. However, with this command:

```
wget -r -l 2 -p http://site/1.html
```

all the above files *and* '3.html''s requisite '3.gif' will be downloaded. Similarly,

```
wget -r -l 1 -p http://site/1.html
```

will cause '1.html', '1.gif', '2.html', and '2.gif' to be downloaded. One might think that:

```
wget -r -l 0 -p http://site/1.html
```

would download just '1.html' and '1.gif', but unfortunately this is not the case, because '-l 0' is equivalent to '-l inf'—that is, infinite recursion. To download a single HTML page (or a handful of them, all specified on the commandline or in a '-i' URL input file) and its (or their) requisites, simply leave off '-r' and '-l':

```
wget -p http://site/1.html
```

Note that Wget will behave as if '-r' had been specified, but only that single page and its requisites will be downloaded. Links from that page to external documents will not be followed. Actually, to download a single page and all its requisites (even if they exist on separate websites), and make sure the lot displays properly locally, this author likes to use a few options in addition to '-p':

```
wget -E -H -k -K -p http://site/document
```

To finish off this topic, it's worth knowing that Wget's idea of an external document link is any URL specified in an <A> tag, an <AREA> tag, or a <LINK> tag other than <LINK REL="stylesheet">.

2.10 Recursive Accept/Reject Options

'-A *acclist* --accept *acclist*'
'-R *rejlist* --reject *rejlist*'

Specify comma-separated lists of file name suffixes or patterns to accept or reject (see Section 4.2 [Types of Files], page 19 for more details).

'-D *domain-list*'
'--domains=*domain-list*'

Set domains to be followed. *domain-list* is a comma-separated list of domains. Note that it does *not* turn on '-H'.

'--exclude-domains *domain-list*'

Specify the domains that are *not* to be followed. (see Section 4.1 [Spanning Hosts], page 18).

'--follow-ftp'

Follow FTP links from HTML documents. Without this option, Wget will ignore all the FTP links.

'--follow-tags=*list*'

Wget has an internal table of HTML tag / attribute pairs that it considers when looking for linked documents during a recursive retrieval. If a user wants only a subset of those tags to be considered, however, he or she should be specify such tags in a comma-separated *list* with this option.

'-G *list*'
'--ignore-tags=*list*'

This is the opposite of the '--follow-tags' option. To skip certain HTML tags when recursively looking for documents to download, specify them in a comma-separated *list*.

In the past, the '-G' option was the best bet for downloading a single page and its requisites, using a commandline like:

```
wget -Ga,area -H -k -K -r http://site/document
```

However, the author of this option came across a page with tags like `<LINK REL="home" HREF="/">` and came to the realization that '-G' was not enough. One can't just tell Wget to ignore `<LINK>`, because then stylesheets will not be downloaded. Now the best bet for downloading a single page and its requisites is the dedicated '--page-requisites' option.

'-H'
'--span-hosts'

Enable spanning across hosts when doing recursive retrieving (see Section 4.1 [Spanning Hosts], page 18).

'-L'
'--relative'

Follow relative links only. Useful for retrieving a specific home page without any distractions, not even those from the same hosts (see Section 4.4 [Relative Links], page 20).

'-I *list*'
'--include-directories=*list*'

Specify a comma-separated list of directories you wish to follow when downloading (see Section 4.3 [Directory-Based Limits], page 19 for more details.) Elements of *list* may contain wildcards.

'-X *list*'
'--exclude-directories=*list*'

Specify a comma-separated list of directories you wish to exclude from download (see Section 4.3 [Directory-Based Limits], page 19 for more details.) Elements of *list* may contain wildcards.

'`-np`'

'`--no-parent`'

Do not ever ascend to the parent directory when retrieving recursively. This is a useful option, since it guarantees that only the files *below* a certain hierarchy will be downloaded. See Section 4.3 [Directory-Based Limits], page 19, for more details.

3 Recursive Retrieval

GNU Wget is capable of traversing parts of the Web (or a single HTTP or FTP server), following links and directory structure. We refer to this as to *recursive retrieving*, or *recursion*.

With HTTP URLs, Wget retrieves and parses the HTML from the given URL, documents, retrieving the files the HTML document was referring to, through markups like `href`, or `src`. If the freshly downloaded file is also of type `text/html`, it will be parsed and followed further.

Recursive retrieval of HTTP and HTML content is *breadth-first*. This means that Wget first downloads the requested HTML document, then the documents linked from that document, then the documents linked by them, and so on. In other words, Wget first downloads the documents at depth 1, then those at depth 2, and so on until the specified maximum depth.

The maximum *depth* to which the retrieval may descend is specified with the '`-l`' option. The default maximum depth is five layers.

When retrieving an FTP URL recursively, Wget will retrieve all the data from the given directory tree (including the subdirectories up to the specified depth) on the remote server, creating its mirror image locally. FTP retrieval is also limited by the `depth` parameter. Unlike HTTP recursion, FTP recursion is performed depth-first.

By default, Wget will create a local directory tree, corresponding to the one found on the remote server.

Recursive retrieving can find a number of applications, the most important of which is mirroring. It is also useful for WWW presentations, and any other opportunities where slow network connections should be bypassed by storing the files locally.

You should be warned that recursive downloads can overload the remote servers. Because of that, many administrators frown upon them and may ban access from your site if they detect very fast downloads of big amounts of content. When downloading from Internet servers, consider using the '`-w`' option to introduce a delay between accesses to the server. The download will take a while longer, but the server administrator will not be alarmed by your rudeness.

Of course, recursive download may cause problems on your machine. If left to run unchecked, it can easily fill up the disk. If downloading from local network, it can also take bandwidth on the system, as well as consume memory and CPU.

Try to specify the criteria that match the kind of download you are trying to achieve. If you want to download only one page, use '`--page-requisites`' without any additional recursion. If you want to download things under one directory, use '`-np`' to avoid downloading things from other directories. If you want to download all the files from one directory, use '`-l 1`' to make sure the recursion depth never exceeds one. See Chapter 4 [Following Links], page 18, for more information about this.

Recursive retrieval should be used with care. Don't say you were not warned.

4 Following Links

When retrieving recursively, one does not wish to retrieve loads of unnecessary data. Most of the time the users bear in mind exactly what they want to download, and want Wget to follow only specific links.

For example, if you wish to download the music archive from '`fly.srk.fer.hr`', you will not want to download all the home pages that happen to be referenced by an obscure part of the archive.

Wget possesses several mechanisms that allows you to fine-tune which links it will follow.

4.1 Spanning Hosts

Wget's recursive retrieval normally refuses to visit hosts different than the one you specified on the command line. This is a reasonable default; without it, every retrieval would have the potential to turn your Wget into a small version of google.

However, visiting different hosts, or *host spanning*, is sometimes a useful option. Maybe the images are served from a different server. Maybe you're mirroring a site that consists of pages interlinked between three servers. Maybe the server has two equivalent names, and the HTML pages refer to both interchangeably.

Span to any host—'`-H`'
: The '`-H`' option turns on host spanning, thus allowing Wget's recursive run to visit any host referenced by a link. Unless sufficient recursion-limiting criteria are applied depth, these foreign hosts will typically link to yet more hosts, and so on until Wget ends up sucking up much more data than you have intended.

Limit spanning to certain domains—'`-D`'
: The '`-D`' option allows you to specify the domains that will be followed, thus limiting the recursion only to the hosts that belong to these domains. Obviously, this makes sense only in conjunction with '`-H`'. A typical example would be downloading the contents of '`www.server.com`', but allowing downloads from '`images.server.com`', etc.:

```
wget -rH -Dserver.com http://www.server.com/
```

You can specify more than one address by separating them with a comma, e.g. '`-Ddomain1.com,domain2.com`'.

Keep download off certain domains—'`--exclude-domains`'
: If there are domains you want to exclude specifically, you can do it with '`--exclude-domains`', which accepts the same type of arguments of '`-D`', but will *exclude* all the listed domains. For example, if you want to download all the hosts from '`foo.edu`' domain, with the exception of '`sunsite.foo.edu`', you can do it like this:

```
wget -rH -Dfoo.edu --exclude-domains sunsite.foo.edu \
    http://www.foo.edu/
```

4.2 Types of Files

When downloading material from the web, you will often want to restrict the retrieval to only certain file types. For example, if you are interested in downloading GIFs, you will not be overjoyed to get loads of PostScript documents, and vice versa.

Wget offers two options to deal with this problem. Each option description lists a short name, a long name, and the equivalent command in '`.wgetrc`'.

'`-A` *acclist*'
'`--accept` *acclist*'
'`accept =` *acclist*'

The argument to '`--accept`' option is a list of file suffixes or patterns that Wget will download during recursive retrieval. A suffix is the ending part of a file, and consists of "normal" letters, e.g. '`gif`' or '`.jpg`'. A matching pattern contains shell-like wildcards, e.g. '`books*`' or '`zelazny*196[0-9]*`'.

So, specifying '`wget -A gif,jpg`' will make Wget download only the files ending with '`gif`' or '`jpg`', i.e. GIFs and JPEGs. On the other hand, '`wget -A "zelazny*196[0-9]*"`' will download only files beginning with '`zelazny`' and containing numbers from 1960 to 1969 anywhere within. Look up the manual of your shell for a description of how pattern matching works.

Of course, any number of suffixes and patterns can be combined into a comma-separated list, and given as an argument to '`-A`'.

'`-R` *rejlist*'
'`--reject` *rejlist*'
'`reject =` *rejlist*'

The '`--reject`' option works the same way as '`--accept`', only its logic is the reverse; Wget will download all files *except* the ones matching the suffixes (or patterns) in the list.

So, if you want to download a whole page except for the cumbersome MPEGs and .AU files, you can use '`wget -R mpg,mpeg,au`'. Analogously, to download all files except the ones beginning with '`bjork`', use '`wget -R "bjork*"`'. The quotes are to prevent expansion by the shell.

The '`-A`' and '`-R`' options may be combined to achieve even better fine-tuning of which files to retrieve. E.g. '`wget -A "*zelazny*" -R .ps`' will download all the files having '`zelazny`' as a part of their name, but *not* the PostScript files.

Note that these two options do not affect the downloading of HTML files; Wget must load all the HTMLs to know where to go at all—recursive retrieval would make no sense otherwise.

4.3 Directory-Based Limits

Regardless of other link-following facilities, it is often useful to place the restriction of what files to retrieve based on the directories those files are placed in. There can be many reasons for this—the home pages may be organized in a reasonable directory structure; or some directories may contain useless information, e.g. '`/cgi-bin`' or '`/dev`' directories.

Wget offers three different options to deal with this requirement. Each option description lists a short name, a long name, and the equivalent command in '`.wgetrc`'.

'`-I` *list*'
'`--include` *list*'
'`include_directories =` *list*'

: '`-I`' option accepts a comma-separated list of directories included in the retrieval. Any other directories will simply be ignored. The directories are absolute paths.

  So, if you wish to download from '`http://host/people/bozo/`' following only links to bozo's colleagues in the '`/people`' directory and the bogus scripts in '`/cgi-bin`', you can specify:

```
wget -I /people,/cgi-bin http://host/people/bozo/
```

'`-X` *list*'
'`--exclude` *list*'
'`exclude_directories =` *list*'

: '`-X`' option is exactly the reverse of '`-I`'—this is a list of directories *excluded* from the download. E.g. if you do not want Wget to download things from '`/cgi-bin`' directory, specify '`-X /cgi-bin`' on the command line.

  The same as with '`-A`'/'`-R`', these two options can be combined to get a better fine-tuning of downloading subdirectories. E.g. if you want to load all the files from '`/pub`' hierarchy except for '`/pub/worthless`', specify '`-I/pub -X/pub/worthless`'.

'`-np`'
'`--no-parent`'
'`no_parent = on`'

: The simplest, and often very useful way of limiting directories is disallowing retrieval of the links that refer to the hierarchy *above* than the beginning directory, i.e. disallowing ascent to the parent directory/directories.

  The '`--no-parent`' option (short '`-np`') is useful in this case. Using it guarantees that you will never leave the existing hierarchy. Supposing you issue Wget with:

```
wget -r --no-parent http://somehost/~luzer/my-archive/
```

  You may rest assured that none of the references to '`/~his-girls-homepage/`' or '`/~luzer/all-my-mpegs/`' will be followed. Only the archive you are interested in will be downloaded. Essentially, '`--no-parent`' is similar to '`-I/~luzer/my-archive`', only it handles redirections in a more intelligent fashion.

4.4 Relative Links

When '`-L`' is turned on, only the relative links are ever followed. Relative links are here defined those that do not refer to the web server root. For example, these links are relative:

```
<a href="foo.gif">
<a href="foo/bar.gif">
<a href="../foo/bar.gif">
```

These links are not relative:

```
<a href="/foo.gif">
<a href="/foo/bar.gif">
```

```
<a href="http://www.server.com/foo/bar.gif">
```

Using this option guarantees that recursive retrieval will not span hosts, even without '-H'. In simple cases it also allows downloads to "just work" without having to convert links.

This option is probably not very useful and might be removed in a future release.

4.5 Following FTP Links

The rules for FTP are somewhat specific, as it is necessary for them to be. FTP links in HTML documents are often included for purposes of reference, and it is often inconvenient to download them by default.

To have FTP links followed from HTML documents, you need to specify the '--follow-ftp' option. Having done that, FTP links will span hosts regardless of '-H' setting. This is logical, as FTP links rarely point to the same host where the HTTP server resides. For similar reasons, the '-L' options has no effect on such downloads. On the other hand, domain acceptance ('-D') and suffix rules ('-A' and '-R') apply normally.

Also note that followed links to FTP directories will not be retrieved recursively further.

5 Time-Stamping

One of the most important aspects of mirroring information from the Internet is updating your archives.

Downloading the whole archive again and again, just to replace a few changed files is expensive, both in terms of wasted bandwidth and money, and the time to do the update. This is why all the mirroring tools offer the option of incremental updating.

Such an updating mechanism means that the remote server is scanned in search of *new* files. Only those new files will be downloaded in the place of the old ones.

A file is considered new if one of these two conditions are met:

1. A file of that name does not already exist locally.
2. A file of that name does exist, but the remote file was modified more recently than the local file.

To implement this, the program needs to be aware of the time of last modification of both local and remote files. We call this information the *time-stamp* of a file.

The time-stamping in GNU Wget is turned on using '`--timestamping`' ('`-N`') option, or through `timestamping = on` directive in '`.wgetrc`'. With this option, for each file it intends to download, Wget will check whether a local file of the same name exists. If it does, and the remote file is older, Wget will not download it.

If the local file does not exist, or the sizes of the files do not match, Wget will download the remote file no matter what the time-stamps say.

5.1 Time-Stamping Usage

The usage of time-stamping is simple. Say you would like to download a file so that it keeps its date of modification.

```
wget -S http://www.gnu.ai.mit.edu/
```

A simple `ls -l` shows that the time stamp on the local file equals the state of the `Last-Modified` header, as returned by the server. As you can see, the time-stamping info is preserved locally, even without '`-N`' (at least for HTTP).

Several days later, you would like Wget to check if the remote file has changed, and download it if it has.

```
wget -N http://www.gnu.ai.mit.edu/
```

Wget will ask the server for the last-modified date. If the local file has the same timestamp as the server, or a newer one, the remote file will not be re-fetched. However, if the remote file is more recent, Wget will proceed to fetch it.

The same goes for FTP. For example:

```
wget "ftp://ftp.ifi.uio.no/pub/emacs/gnus/*"
```

(The quotes around that URL are to prevent the shell from trying to interpret the '`*`'.)

After download, a local directory listing will show that the timestamps match those on the remote server. Reissuing the command with '`-N`' will make Wget re-fetch *only* the files that have been modified since the last download.

If you wished to mirror the GNU archive every week, you would use a command like the following, weekly:

```
wget --timestamping -r ftp://ftp.gnu.org/pub/gnu/
```

Note that time-stamping will only work for files for which the server gives a timestamp. For HTTP, this depends on getting a `Last-Modified` header. For FTP, this depends on getting a directory listing with dates in a format that Wget can parse (see Section 5.3 [FTP Time-Stamping Internals], page 23).

5.2 HTTP Time-Stamping Internals

Time-stamping in HTTP is implemented by checking of the `Last-Modified` header. If you wish to retrieve the file '`foo.html`' through HTTP, Wget will check whether '`foo.html`' exists locally. If it doesn't, '`foo.html`' will be retrieved unconditionally.

If the file does exist locally, Wget will first check its local time-stamp (similar to the way `ls -l` checks it), and then send a `HEAD` request to the remote server, demanding the information on the remote file.

The `Last-Modified` header is examined to find which file was modified more recently (which makes it "newer"). If the remote file is newer, it will be downloaded; if it is older, Wget will give up.[1]

When '`--backup-converted`' ('`-K`') is specified in conjunction with '`-N`', server file '*X*' is compared to local file '*X*`.orig`', if extant, rather than being compared to local file '*X*', which will always differ if it's been converted by '`--convert-links`' ('`-k`').

Arguably, HTTP time-stamping should be implemented using the `If-Modified-Since` request.

5.3 FTP Time-Stamping Internals

In theory, FTP time-stamping works much the same as HTTP, only FTP has no headers—time-stamps must be ferreted out of directory listings.

[1] As an additional check, Wget will look at the `Content-Length` header, and compare the sizes; if they are not the same, the remote file will be downloaded no matter what the time-stamp says.

If an FTP download is recursive or uses globbing, Wget will use the FTP `LIST` command to get a file listing for the directory containing the desired file(s). It will try to analyze the listing, treating it like Unix `ls -l` output, extracting the time-stamps. The rest is exactly the same as for HTTP. Note that when retrieving individual files from an FTP server without using globbing or recursion, listing files will not be downloaded (and thus files will not be time-stamped) unless '`-N`' is specified.

Assumption that every directory listing is a Unix-style listing may sound extremely constraining, but in practice it is not, as many non-Unix FTP servers use the Unixoid listing format because most (all?) of the clients understand it. Bear in mind that RFC959 defines no standard way to get a file list, let alone the time-stamps. We can only hope that a future standard will define this.

Another non-standard solution includes the use of `MDTM` command that is supported by some FTP servers (including the popular `wu-ftpd`), which returns the exact time of the specified file. Wget may support this command in the future.

6 Startup File

Once you know how to change default settings of Wget through command line arguments, you may wish to make some of those settings permanent. You can do that in a convenient way by creating the Wget startup file—'.wgetrc'.

Besides '.wgetrc' is the "main" initialization file, it is convenient to have a special facility for storing passwords. Thus Wget reads and interprets the contents of '$HOME/.netrc', if it finds it. You can find '.netrc' format in your system manuals.

Wget reads '.wgetrc' upon startup, recognizing a limited set of commands.

6.1 Wgetrc Location

When initializing, Wget will look for a *global* startup file, '/usr/local/etc/wgetrc' by default (or some prefix other than '/usr/local', if Wget was not installed there) and read commands from there, if it exists.

Then it will look for the user's file. If the environmental variable WGETRC is set, Wget will try to load that file. Failing that, no further attempts will be made.

If WGETRC is not set, Wget will try to load '$HOME/.wgetrc'.

The fact that user's settings are loaded after the system-wide ones means that in case of collision user's wgetrc *overrides* the system-wide wgetrc (in '/usr/local/etc/wgetrc' by default). Fascist admins, away!

6.2 Wgetrc Syntax

The syntax of a wgetrc command is simple:

```
variable = value
```

The *variable* will also be called *command*. Valid *values* are different for different commands.

The commands are case-insensitive and underscore-insensitive. Thus 'DIr__PrefiX' is the same as 'dirprefix'. Empty lines, lines beginning with '#' and lines containing white-space only are discarded.

Commands that expect a comma-separated list will clear the list on an empty command. So, if you wish to reset the rejection list specified in global 'wgetrc', you can do it with:

```
reject =
```

6.3 Wgetrc Commands

The complete set of commands is listed below. Legal values are listed after the '`=`'. Simple Boolean values can be set or unset using '`on`' and '`off`' or '`1`' and '`0`'. A fancier kind of Boolean allowed in some cases is the *lockable Boolean*, which may be set to '`on`', '`off`', '`always`', or '`never`'. If an option is set to '`always`' or '`never`', that value will be locked in for the duration of the Wget invocation—commandline options will not override.

Some commands take pseudo-arbitrary values. *address* values can be hostnames or dotted-quad IP addresses. *n* can be any positive integer, or '`inf`' for infinity, where appropriate. *string* values can be any non-empty string.

Most of these commands have commandline equivalents (see Chapter 2 [Invoking], page 2), though some of the more obscure or rarely used ones do not.

accept/reject = *string*
: Same as '`-A`'/'`-R`' (see Section 4.2 [Types of Files], page 19).

add_hostdir = on/off
: Enable/disable host-prefixed file names. '`-nH`' disables it.

continue = on/off
: If set to on, force continuation of preexistent partially retrieved files. See '`-c`' before setting it.

background = on/off
: Enable/disable going to background—the same as '`-b`' (which enables it).

backup_converted = on/off
: Enable/disable saving pre-converted files with the suffix '`.orig`'—the same as '`-K`' (which enables it).

base = *string*
: Consider relative URLs in URL input files forced to be interpreted as HTML as being relative to *string*—the same as '`-B`'.

bind_address = *address*
: Bind to *address*, like the '`--bind-address`' option.

cache = on/off
: When set to off, disallow server-caching. See the '`-C`' option.

convert links = on/off
: Convert non-relative links locally. The same as '`-k`'.

cookies = on/off
: When set to off, disallow cookies. See the '`--cookies`' option.

load_cookies = *file*
: Load cookies from *file*. See '`--load-cookies`'.

save_cookies = *file*
: Save cookies to *file*. See '`--save-cookies`'.

cut_dirs = *n*
: Ignore *n* remote directory components.

debug = on/off
: Debug mode, same as '`-d`'.

delete_after = on/off
: Delete after download—the same as '--delete-after'.

dir_prefix = *string*
: Top of directory tree—the same as '-P'.

dirstruct = on/off
: Turning dirstruct on or off—the same as '-x' or '-nd', respectively.

domains = *string*
: Same as '-D' (see Section 4.1 [Spanning Hosts], page 18).

dot_bytes = *n*
: Specify the number of bytes "contained" in a dot, as seen throughout the retrieval (1024 by default). You can postfix the value with 'k' or 'm', representing kilobytes and megabytes, respectively. With dot settings you can tailor the dot retrieval to suit your needs, or you can use the predefined *styles* (see Section 2.5 [Download Options], page 5).

dots_in_line = *n*
: Specify the number of dots that will be printed in each line throughout the retrieval (50 by default).

dot_spacing = *n*
: Specify the number of dots in a single cluster (10 by default).

exclude_directories = *string*
: Specify a comma-separated list of directories you wish to exclude from download—the same as '-X' (see Section 4.3 [Directory-Based Limits], page 19).

exclude_domains = *string*
: Same as '--exclude-domains' (see Section 4.1 [Spanning Hosts], page 18).

follow_ftp = on/off
: Follow FTP links from HTML documents—the same as '--follow-ftp'.

follow_tags = *string*
: Only follow certain HTML tags when doing a recursive retrieval, just like '--follow-tags'.

force_html = on/off
: If set to on, force the input filename to be regarded as an HTML document—the same as '-F'.

ftp_proxy = *string*
: Use *string* as FTP proxy, instead of the one specified in environment.

glob = on/off
: Turn globbing on/off—the same as '-g'.

header = *string*
: Define an additional header, like '--header'.

html_extension = on/off
: Add a '.html' extension to 'text/html' files without it, like '-E'.

http_passwd = *string*
: Set HTTP password.

http_proxy = *string*
: Use *string* as HTTP proxy, instead of the one specified in environment.

http_user = *string*
: Set HTTP user to *string*.

ignore_length = on/off
    When set to on, ignore `Content-Length` header; the same as '`--ignore-length`'.

ignore_tags = *string*
    Ignore certain HTML tags when doing a recursive retrieval, just like '`-G`' / '`--ignore-tags`'.

include_directories = *string*
    Specify a comma-separated list of directories you wish to follow when downloading—the same as '`-I`'.

input = *string*
    Read the URLs from *string*, like '`-i`'.

kill_longer = on/off
    Consider data longer than specified in content-length header as invalid (and retry getting it). The default behaviour is to save as much data as there is, provided there is more than or equal to the value in `Content-Length`.

logfile = *string*
    Set logfile—the same as '`-o`'.

login = *string*
    Your user name on the remote machine, for FTP. Defaults to '`anonymous`'.

mirror = on/off
    Turn mirroring on/off. The same as '`-m`'.

netrc = on/off
    Turn reading netrc on or off.

noclobber = on/off
    Same as '`-nc`'.

no_parent = on/off
    Disallow retrieving outside the directory hierarchy, like '`--no-parent`' (see Section 4.3 [Directory-Based Limits], page 19).

no_proxy = *string*
    Use *string* as the comma-separated list of domains to avoid in proxy loading, instead of the one specified in environment.

output_document = *string*
    Set the output filename—the same as '`-O`'.

page_requisites = on/off
    Download all ancillary documents necessary for a single HTML page to display properly—the same as '`-p`'.

passive_ftp = on/off/always/never
    Set passive FTP—the same as '`--passive-ftp`'. Some scripts and '`.pm`' (Perl module) files download files using '`wget --passive-ftp`'. If your firewall does not allow this, you can set '`passive_ftp = never`' to override the commandline.

passwd = *string*
    Set your FTP password to *password*. Without this setting, the password defaults to '`username@hostname.domainname`'.

progress = *string*
    Set the type of the progress indicator. Legal types are "dot" and "bar".

proxy_user = *string*
    Set proxy authentication user name to *string*, like '`--proxy-user`'.

proxy_passwd = *string*
: Set proxy authentication password to *string*, like '`--proxy-passwd`'.

referer = *string*
: Set HTTP '`Referer:`' header just like '`--referer`'. (Note it was the folks who wrote the HTTP spec who got the spelling of "referrer" wrong.)

quiet = on/off
: Quiet mode—the same as '`-q`'.

quota = *quota*
: Specify the download quota, which is useful to put in the global '`wgetrc`'. When download quota is specified, Wget will stop retrieving after the download sum has become greater than quota. The quota can be specified in bytes (default), kbytes '`k`' appended) or mbytes ('`m`' appended). Thus '`quota = 5m`' will set the quota to 5 mbytes. Note that the user's startup file overrides system settings.

reclevel = *n*
: Recursion level—the same as '`-l`'.

recursive = on/off
: Recursive on/off—the same as '`-r`'.

relative_only = on/off
: Follow only relative links—the same as '`-L`' (see Section 4.4 [Relative Links], page 20).

remove_listing = on/off
: If set to on, remove FTP listings downloaded by Wget. Setting it to off is the same as '`-nr`'.

retr_symlinks = on/off
: When set to on, retrieve symbolic links as if they were plain files; the same as '`--retr-symlinks`'.

robots = on/off
: Use (or not) '`/robots.txt`' file (see Section 9.1 [Robots], page 39). Be sure to know what you are doing before changing the default (which is '`on`').

server_response = on/off
: Choose whether or not to print the HTTP and FTP server responses—the same as '`-S`'.

span_hosts = on/off
: Same as '`-H`'.

timeout = *n*
: Set timeout value—the same as '`-T`'.

timestamping = on/off
: Turn timestamping on/off. The same as '`-N`' (see Chapter 5 [Time-Stamping], page 22).

tries = *n*
: Set number of retries per URL—the same as '`-t`'.

use_proxy = on/off
: Turn proxy support on/off. The same as '`-Y`'.

verbose = on/off
: Turn verbose on/off—the same as '`-v`'/'`-nv`'.

wait = *n*
: Wait *n* seconds between retrievals—the same as '`-w`'.

waitretry = *n*
: Wait up to *n* seconds between retries of failed retrievals only—the same as '`--waitretry`'. Note that this is turned on by default in the global '`wgetrc`'.

randomwait = on/off
: Turn random between-request wait times on or off. The same as '`--random-wait`'.

6.4 Sample Wgetrc

This is the sample initialization file, as given in the distribution. It is divided in two section—one for global usage (suitable for global startup file), and one for local usage (suitable for '$HOME/.wgetrc'). Be careful about the things you change.

Note that almost all the lines are commented out. For a command to have any effect, you must remove the '#' character at the beginning of its line.

```
###
### Sample Wget initialization file .wgetrc
###

## You can use this file to change the default behaviour of wget or to
## avoid having to type many many command-line options. This file does
## not contain a comprehensive list of commands -- look at the manual
## to find out what you can put into this file.
##
## Wget initialization file can reside in /usr/local/etc/wgetrc
## (global, for all users) or $HOME/.wgetrc (for a single user).
##
## To use the settings in this file, you will have to uncomment them,
## as well as change them, in most cases, as the values on the
## commented-out lines are the default values (e.g. "off").


##
## Global settings (useful for setting up in /usr/local/etc/wgetrc).
## Think well before you change them, since they may reduce wget's
## functionality, and make it behave contrary to the documentation:
##

# You can set retrieve quota for beginners by specifying a value
# optionally followed by 'K' (kilobytes) or 'M' (megabytes).  The
# default quota is unlimited.
#quota = inf

# You can lower (or raise) the default number of retries when
# downloading a file (default is 20).
#tries = 20

# Lowering the maximum depth of the recursive retrieval is handy to
# prevent newbies from going too "deep" when they unwittingly start
# the recursive retrieval.  The default is 5.
#reclevel = 5

# Many sites are behind firewalls that do not allow initiation of
# connections from the outside.  On these sites you have to use the
# `passive' feature of FTP.  If you are behind such a firewall, you
# can turn this on to make Wget use passive FTP by default.
#passive_ftp = off
```

```
# The "wait" command below makes Wget wait between every connection.
# If, instead, you want Wget to wait only between retries of failed
# downloads, set waitretry to maximum number of seconds to wait (Wget
# will use "linear backoff", waiting 1 second after the first failure
# on a file, 2 seconds after the second failure, etc. up to this max).
waitretry = 10


##
## Local settings (for a user to set in his $HOME/.wgetrc).  It is
## *highly* undesirable to put these settings in the global file, since
## they are potentially dangerous to "normal" users.
##
## Even when setting up your own ~/.wgetrc, you should know what you
## are doing before doing so.
##

# Set this to on to use timestamping by default:
#timestamping = off

# It is a good idea to make Wget send your email address in a `From:'
# header with your request (so that server administrators can contact
# you in case of errors).  Wget does *not* send `From:' by default.
#header = From: Your Name <username@site.domain>

# You can set up other headers, like Accept-Language.  Accept-Language
# is *not* sent by default.
#header = Accept-Language: en

# You can set the default proxies for Wget to use for http and ftp.
# They will override the value in the environment.
#http_proxy = http://proxy.yoyodyne.com:18023/
#ftp_proxy = http://proxy.yoyodyne.com:18023/

# If you do not want to use proxy at all, set this to off.
#use_proxy = on

# You can customize the retrieval outlook.  Valid options are default,
# binary, mega and micro.
#dot_style = default

# Setting this to off makes Wget not download /robots.txt.  Be sure to
# know *exactly* what /robots.txt is and how it is used before changing
# the default!
#robots = on

# It can be useful to make Wget wait between connections.  Set this to
# the number of seconds you want Wget to wait.
#wait = 0
```

```
# You can force creating directory structure, even if a single is being
# retrieved, by setting this to on.
#dirstruct = off

# You can turn on recursive retrieving by default (don't do this if
# you are not sure you know what it means) by setting this to on.
#recursive = off

# To always back up file X as X.orig before converting its links (due
# to -k / --convert-links / convert_links = on having been specified),
# set this variable to on:
#backup_converted = off

# To have Wget follow FTP links from HTML files by default, set this
# to on:
#follow_ftp = off
```

7 Examples

The examples are divided into three sections loosely based on their complexity.

7.1 Simple Usage

- Say you want to download a URL. Just type:

  ```
  wget http://fly.srk.fer.hr/
  ```

- But what will happen if the connection is slow, and the file is lengthy? The connection will probably fail before the whole file is retrieved, more than once. In this case, Wget will try getting the file until it either gets the whole of it, or exceeds the default number of retries (this being 20). It is easy to change the number of tries to 45, to insure that the whole file will arrive safely:

  ```
  wget --tries=45 http://fly.srk.fer.hr/jpg/flyweb.jpg
  ```

- Now let's leave Wget to work in the background, and write its progress to log file '`log`'. It is tiring to type '`--tries`', so we shall use '`-t`'.

  ```
  wget -t 45 -o log http://fly.srk.fer.hr/jpg/flyweb.jpg &
  ```

  The ampersand at the end of the line makes sure that Wget works in the background. To unlimit the number of retries, use '`-t inf`'.

- The usage of FTP is as simple. Wget will take care of login and password.

  ```
  wget ftp://gnjilux.srk.fer.hr/welcome.msg
  ```

- If you specify a directory, Wget will retrieve the directory listing, parse it and convert it to HTML. Try:

  ```
  wget ftp://prep.ai.mit.edu/pub/gnu/
  links index.html
  ```

7.2 Advanced Usage

- You have a file that contains the URLs you want to download? Use the '`-i`' switch:

  `wget -i` *file*

  If you specify '`-`' as file name, the URLs will be read from standard input.

- Create a five levels deep mirror image of the GNU web site, with the same directory structure the original has, with only one try per document, saving the log of the activities to '`gnulog`':

  ```
  wget -r http://www.gnu.org/ -o gnulog
  ```

- The same as the above, but convert the links in the HTML files to point to local files, so you can view the documents off-line:

  ```
  wget --convert-links -r http://www.gnu.org/ -o gnulog
  ```

- Retrieve only one HTML page, but make sure that all the elements needed for the page to be displayed, such as inline images and external style sheets, are also downloaded. Also make sure the downloaded page references the downloaded links.

  ```
  wget -p --convert-links http://www.server.com/dir/page.html
  ```

  The HTML page will be saved to '`www.server.com/dir/page.html`', and the images, stylesheets, etc., somewhere under '`www.server.com/`', depending on where they were on the remote server.

- The same as the above, but without the 'www.server.com/' directory. In fact, I don't want to have all those random server directories anyway—just save *all* those files under a 'download/' subdirectory of the current directory.

```
wget -p --convert-links -nH -nd -Pdownload \
     http://www.server.com/dir/page.html
```

- Retrieve the index.html of 'www.lycos.com', showing the original server headers:

```
wget -S http://www.lycos.com/
```

- Save the server headers with the file, perhaps for post-processing.

```
wget -s http://www.lycos.com/
more index.html
```

- Retrieve the first two levels of 'wuarchive.wustl.edu', saving them to '/tmp'.

```
wget -r -l2 -P/tmp ftp://wuarchive.wustl.edu/
```

- You want to download all the GIFs from a directory on an HTTP server. You tried 'wget http://www.server.com/dir/*.gif', but that didn't work because HTTP retrieval does not support globbing. In that case, use:

```
wget -r -l1 --no-parent -A.gif http://www.server.com/dir/
```

  More verbose, but the effect is the same. '-r -l1' means to retrieve recursively (see Chapter 3 [Recursive Retrieval], page 17), with maximum depth of 1. '--no-parent' means that references to the parent directory are ignored (see Section 4.3 [Directory-Based Limits], page 19), and '-A.gif' means to download only the GIF files. '-A "*.gif"' would have worked too.

- Suppose you were in the middle of downloading, when Wget was interrupted. Now you do not want to clobber the files already present. It would be:

```
wget -nc -r http://www.gnu.org/
```

- If you want to encode your own username and password to HTTP or FTP, use the appropriate URL syntax (see Section 2.1 [URL Format], page 2).

```
wget ftp://hniksic:mypassword@unix.server.com/.emacs
```

- You would like the output documents to go to standard output instead of to files?

```
wget -O - http://jagor.srce.hr/ http://www.srce.hr/
```

  You can also combine the two options and make pipelines to retrieve the documents from remote hotlists:

```
wget -O - http://cool.list.com/ | wget --force-html -i -
```

7.3 Very Advanced Usage

- If you wish Wget to keep a mirror of a page (or FTP subdirectories), use '--mirror' ('-m'), which is the shorthand for '-r -l inf -N'. You can put Wget in the crontab file asking it to recheck a site each Sunday:

```
crontab
0 0 * * 0 wget --mirror http://www.gnu.org/ -o /home/me/weeklog
```

- In addition to the above, you want the links to be converted for local viewing. But, after having read this manual, you know that link conversion doesn't play well with timestamping, so you also want Wget to back up the original HTML files before the conversion. Wget invocation would look like this:

```
wget --mirror --convert-links --backup-converted  \
     http://www.gnu.org/ -o /home/me/weeklog
```

- But you've also noticed that local viewing doesn't work all that well when HTML files are saved under extensions other than '`.html`', perhaps because they were served as '`index.cgi`'. So you'd like Wget to rename all the files served with content-type '`text/html`' to '*name*`.html`'.

  ```
  wget --mirror --convert-links --backup-converted \
       --html-extension -o /home/me/weeklog        \
       http://www.gnu.org/
  ```

  Or, with less typing:

  ```
  wget -m -k -K -E http://www.gnu.org/ -o /home/me/weeklog
  ```

8 Various

This chapter contains all the stuff that could not fit anywhere else.

8.1 Proxies

Proxies are special-purpose HTTP servers designed to transfer data from remote servers to local clients. One typical use of proxies is lightening network load for users behind a slow connection. This is achieved by channeling all HTTP and FTP requests through the proxy which caches the transferred data. When a cached resource is requested again, proxy will return the data from cache. Another use for proxies is for companies that separate (for security reasons) their internal networks from the rest of Internet. In order to obtain information from the Web, their users connect and retrieve remote data using an authorized proxy.

Wget supports proxies for both HTTP and FTP retrievals. The standard way to specify proxy location, which Wget recognizes, is using the following environment variables:

`http_proxy`
: This variable should contain the URL of the proxy for HTTP connections.

`ftp_proxy`
: This variable should contain the URL of the proxy for FTP connections. It is quite common that HTTP_PROXY and FTP_PROXY are set to the same URL.

`no_proxy`
: This variable should contain a comma-separated list of domain extensions proxy should *not* be used for. For instance, if the value of `no_proxy` is '`.mit.edu`', proxy will not be used to retrieve documents from MIT.

In addition to the environment variables, proxy location and settings may be specified from within Wget itself.

'`-Y on/off`'
'`--proxy=on/off`'
'`proxy = on/off`'
: This option may be used to turn the proxy support on or off. Proxy support is on by default, provided that the appropriate environment variables are set.

'`http_proxy =` *URL*'
'`ftp_proxy =` *URL*'
'`no_proxy =` *string*'
: These startup file variables allow you to override the proxy settings specified by the environment.

Some proxy servers require authorization to enable you to use them. The authorization consists of *username* and *password*, which must be sent by Wget. As with HTTP authorization, several authentication schemes exist. For proxy authorization only the `Basic` authentication scheme is currently implemented.

You may specify your username and password either through the proxy URL or through the command-line options. Assuming that the company's proxy is located at '`proxy.company.com`' at port 8001, a proxy URL location containing authorization data might look like this:

```
http://hniksic:mypassword@proxy.company.com:8001/
```

Alternatively, you may use the 'proxy-user' and 'proxy-password' options, and the equivalent '.wgetrc' settings proxy_user and proxy_passwd to set the proxy username and password.

8.2 Distribution

Like all GNU utilities, the latest version of Wget can be found at the master GNU archive site prep.ai.mit.edu, and its mirrors. For example, Wget 1.8.1 can be found at ftp://prep.ai.mit.edu/gnu/wget/wget-1.8.1.tar.gz

8.3 Mailing List

Wget has its own mailing list at wget@sunsite.dk, thanks to Karsten Thygesen. The mailing list is for discussion of Wget features and web, reporting Wget bugs (those that you think may be of interest to the public) and mailing announcements. You are welcome to subscribe. The more people on the list, the better!

To subscribe, send mail to wget-subscribe@sunsite.dk. the magic word 'subscribe' in the subject line. Unsubscribe by mailing to wget-unsubscribe@sunsite.dk.

The mailing list is archived at http://fly.srk.fer.hr/archive/wget. Alternative archive is available at http://www.mail-archive.com/wget%40sunsite.auc.dk/.

8.4 Reporting Bugs

You are welcome to send bug reports about GNU Wget to bug-wget@gnu.org.

Before actually submitting a bug report, please try to follow a few simple guidelines.

1. Please try to ascertain that the behaviour you see really is a bug. If Wget crashes, it's a bug. If Wget does not behave as documented, it's a bug. If things work strange, but you are not sure about the way they are supposed to work, it might well be a bug.
2. Try to repeat the bug in as simple circumstances as possible. E.g. if Wget crashes while downloading 'wget -rl0 -kKE -t5 -Y0 http://yoyodyne.com -o /tmp/log', you should try to see if the crash is repeatable, and if will occur with a simpler set of options. You might even try to start the download at the page where the crash occurred to see if that page somehow triggered the crash.

   Also, while I will probably be interested to know the contents of your '.wgetrc' file, just dumping it into the debug message is probably a bad idea. Instead, you should first try to see if the bug repeats with '.wgetrc' moved out of the way. Only if it turns out that '.wgetrc' settings affect the bug, mail me the relevant parts of the file.
3. Please start Wget with '-d' option and send the log (or the relevant parts of it). If Wget was compiled without debug support, recompile it. It is *much* easier to trace bugs with debug support on.

4. If Wget has crashed, try to run it in a debugger, e.g. `gdb `which wget` core` and type `where` to get the backtrace.

8.5 Portability

Since Wget uses GNU Autoconf for building and configuring, and avoids using "special" ultra–mega–cool features of any particular Unix, it should compile (and work) on all common Unix flavors.

Various Wget versions have been compiled and tested under many kinds of Unix systems, including Solaris, Linux, SunOS, OSF (aka Digital Unix), Ultrix, *BSD, IRIX, and others; refer to the file 'MACHINES' in the distribution directory for a comprehensive list. If you compile it on an architecture not listed there, please let me know so I can update it.

Wget should also compile on the other Unix systems, not listed in 'MACHINES'. If it doesn't, please let me know.

Thanks to kind contributors, this version of Wget compiles and works on Microsoft Windows 95 and Windows NT platforms. It has been compiled successfully using MS Visual C++ 4.0, Watcom, and Borland C compilers, with Winsock as networking software. Naturally, it is crippled of some features available on Unix, but it should work as a substitute for people stuck with Windows. Note that the Windows port is **neither tested nor maintained** by me—all questions and problems should be reported to Wget mailing list at wget@sunsite.dk where the maintainers will look at them.

8.6 Signals

Since the purpose of Wget is background work, it catches the hangup signal (SIGHUP) and ignores it. If the output was on standard output, it will be redirected to a file named 'wget-log'. Otherwise, SIGHUP is ignored. This is convenient when you wish to redirect the output of Wget after having started it.

```
$ wget http://www.ifi.uio.no/~larsi/gnus.tar.gz &
$ kill -HUP %%     # Redirect the output to wget-log
```

Other than that, Wget will not try to interfere with signals in any way. *C-c*, `kill -TERM` and `kill -KILL` should kill it alike.

9 Appendices

This chapter contains some references I consider useful.

9.1 Robots

It is extremely easy to make Wget wander aimlessly around a web site, sucking all the available data in progress. '`wget -r` *site*', and you're set. Great? Not for the server admin.

While Wget is retrieving static pages, there's not much of a problem. But for Wget, there is no real difference between a static page and the most demanding CGI. For instance, a site I know has a section handled by an, uh, *bitchin'* CGI script that converts all the Info files to HTML. The script can and does bring the machine to its knees without providing anything useful to the downloader.

For such and similar cases various robot exclusion schemes have been devised as a means for the server administrators and document authors to protect chosen portions of their sites from the wandering of robots.

The more popular mechanism is the *Robots Exclusion Standard*, or RES, written by Martijn Koster et al. in 1994. It specifies the format of a text file containing directives that instruct the robots which URL paths to avoid. To be found by the robots, the specifications must be placed in '`/robots.txt`' in the server root, which the robots are supposed to download and parse.

Wget supports RES when downloading recursively. So, when you issue:

```
wget -r http://www.server.com/
```

First the index of '`www.server.com`' will be downloaded. If Wget finds that it wants to download more documents from that server, it will request '`http://www.server.com/robots.txt`' and, if found, use it for further downloads. '`robots.txt`' is loaded only once per each server.

Until version 1.8, Wget supported the first version of the standard, written by Martijn Koster in 1994 and available at `http://www.robotstxt.org/wc/norobots.html`. As of version 1.8, Wget has supported the additional directives specified in the internet draft '`<draft-koster-robots-00.txt>`' titled "A Method for Web Robots Control". The draft, which has as far as I know never made to an RFC, is available at `http://www.robotstxt.org/wc/norobots-rfc.txt`.

This manual no longer includes the text of the Robot Exclusion Standard.

The second, less known mechanism, enables the author of an individual document to specify whether they want the links from the file to be followed by a robot. This is achieved using the `META` tag, like this:

```
<meta name="robots" content="nofollow">
```

This is explained in some detail at `http://www.robotstxt.org/wc/meta-user.html`. Wget supports this method of robot exclusion in addition to the usual '`/robots.txt`' exclusion.

9.2 Security Considerations

When using Wget, you must be aware that it sends unencrypted passwords through the network, which may present a security problem. Here are the main issues, and some solutions.

1. The passwords on the command line are visible using `ps`. If this is a problem, avoid putting passwords from the command line—e.g. you can use '`.netrc`' for this.
2. Using the insecure *basic* authentication scheme, unencrypted passwords are transmitted through the network routers and gateways.
3. The FTP passwords are also in no way encrypted. There is no good solution for this at the moment.
4. Although the "normal" output of Wget tries to hide the passwords, debugging logs show them, in all forms. This problem is avoided by being careful when you send debug logs (yes, even when you send them to me).

9.3 Contributors

GNU Wget was written by Hrvoje Nikšić `hniksic@arsdigita.com`. However, its development could never have gone as far as it has, were it not for the help of many people, either with bug reports, feature proposals, patches, or letters saying "Thanks!".

Special thanks goes to the following people (no particular order):

- Karsten Thygesen—donated system resources such as the mailing list, web space, and FTP space, along with a lot of time to make these actually work.
- Shawn McHorse—bug reports and patches.
- Kaveh R. Ghazi—on-the-fly `ansi2knr`-ization. Lots of portability fixes.
- Gordon Matzigkeit—'`.netrc`' support.
- Zlatko Čalušić, Tomislav Vujec and Dražen Kačar—feature suggestions and "philosophical" discussions.
- Darko Budor—initial port to Windows.
- Antonio Rosella—help and suggestions, plus the Italian translation.
- Tomislav Petrović, Mario Mikočević—many bug reports and suggestions.
- François Pinard—many thorough bug reports and discussions.
- Karl Eichwalder—lots of help with internationalization and other things.
- Junio Hamano—donated support for Opie and HTTP `Digest` authentication.
- The people who provided donations for development, including Brian Gough.

The following people have provided patches, bug/build reports, useful suggestions, beta testing services, fan mail and all the other things that make maintenance so much fun:

Ian Abbott Tim Adam, Adrian Aichner, Martin Baehr, Dieter Baron, Roger Beeman, Dan Berger, T. Bharath, Paul Bludov, Daniel Bodea, Mark Boyns, John Burden, Wanderlei Cavassin, Gilles Cedoc, Tim Charron, Noel Cragg, Kristijan Čonkaš, John Daily, Andrew Davison, Andrew Deryabin, Ulrich Drepper, Marc Duponcheel, Damir Džeko, Alan Eldridge, Aleksandar Erkalović, Andy Eskilsson, Christian Fraenkel, Masashi Fujita, Howard Gayle, Marcel Gerrits, Lemble Gregory, Hans Grobler, Mathieu Guillaume, Dan Harkless, Herold Heiko, Jochen Hein, Karl Heuer, HIROSE Masaaki, Gregor Hoffleit, Erik Magnus Hulthen, Richard Huveneers, Jonas Jensen, Simon Josefsson, Mario Jurić, Hack Kampbjørn, Const Kaplinsky, Goran Kezunović, Robert Kleine, KOJIMA Haime, Fila Kolodny, Alexander Kourakos, Martin Kraemer, *Σίμος Ξενιτέλλης* (Simos KSenitellis), Hrvoje Lacko, Daniel S. Lewart, Nicolás Lichtmeier, Dave Love, Alexander V. Lukyanov, Jordan Mendelson, Lin Zhe Min, Tim Mooney, Simon Munton, Charlie Negyesi, R. K. Owen, Andrew Pollock, Steve Pothier, Jan Přikryl, Marin Purgar, Csaba Ráduly, Keith Refson, Tyler Riddle, Tobias Ringstrom, Juan José Rodrígues, Edward J. Sabol, Heinz Salzmann, Robert Schmidt, Andreas Schwab, Chris Seawood, Toomas Soome, Tage Stabell-Kulo, Sven Sternberger, Markus Strasser, John Summerfield, Szakacsits Szabolcs, Mike Thomas, Philipp Thomas, Dave Turner, Russell Vincent, Charles G Waldman, Douglas E. Wegscheid, Jasmin Zainul, Bojan Ždrnja, Kristijan Zimmer.

Apologies to all who I accidentally left out, and many thanks to all the subscribers of the Wget mailing list.

10 Copying

GNU Wget is licensed under the GNU GPL, which makes it *free software*.

Please note that "free" in "free software" refers to liberty, not price. As some GNU project advocates like to point out, think of "free speech" rather than "free beer". The exact and legally binding distribution terms are spelled out below; in short, you have the right (freedom) to run and change Wget and distribute it to other people, and even—if you want—charge money for doing either. The important restriction is that you have to grant your recipients the same rights and impose the same restrictions.

This method of licensing software is also known as *open source* because, among other things, it makes sure that all recipients will receive the source code along with the program, and be able to improve it. The GNU project prefers the term "free software" for reasons outlined at `http://www.gnu.org/philosophy/free-software-for-freedom.html`.

The exact license terms are defined by this paragraph and the GNU General Public License it refers to:

> GNU Wget is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.
>
> GNU Wget is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.
>
> A copy of the GNU General Public License is included as part of this manual; if you did not receive it, write to the Free Software Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.

In addition to this, this manual is free in the same sense:

> Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with the Invariant Sections being "GNU General Public License" and "GNU Free Documentation License", with no Front-Cover Texts, and with no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

The full texts of the GNU General Public License and of the GNU Free Documentation License are available below.

10.1 GNU General Public License

Version 2, June 1991

675 Mass Ave, Cambridge, MA 02139, USA

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software—to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

1. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License.

The “Program”, below, refers to any such program or work, and a “work based on the Program” means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term “modification”.) Each licensee is addressed as “you”.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

2. You may copy and distribute verbatim copies of the Program’s source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

   You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

3. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

   a. You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

   b. You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

   c. If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

   These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

   Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

   In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

4. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

   a. Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
   b. Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,
   c. Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

   The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

   If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.
5. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.
6. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.
7. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.
8. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

   If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

   It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

9. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.
10. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

    Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.
11. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

12. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.
13. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

```
one line to give the program's name and an idea of what it does.
Copyright (C) 19yy  name of author

This program is free software; you can redistribute it and/or
modify it under the terms of the GNU General Public License
as published by the Free Software Foundation; either version 2
of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program; if not, write to the Free Software
Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.
```

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

```
Gnomovision version 69, Copyright (C) 19yy name of author
Gnomovision comes with ABSOLUTELY NO WARRANTY; for details
type 'show w'.  This is free software, and you are welcome
to redistribute it under certain conditions; type 'show c'
for details.
```

The hypothetical commands '`show w`' and '`show c`' should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than '`show w`' and '`show c`'; they could even be mouse-clicks or menu items—whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the program, if necessary. Here is a sample; alter the names:

```
Yoyodyne, Inc., hereby disclaims all copyright
interest in the program 'Gnomovision'
(which makes passes at compilers) written
by James Hacker.

signature of Ty Coon, 1 April 1989
Ty Coon, President of Vice
```

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.

10.2 GNU Free Documentation License

Version 1.1, March 2000

0. PREAMBLE

   The purpose of this License is to make a manual, textbook, or other written document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

   This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

   We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

   This License applies to any manual or other work that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you".

   A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

   A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (For example, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, whose contents can be viewed and edited directly and straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup has been designed to thwart or discourage subsequent modification by readers is not Transparent. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML designed for human modification. Opaque formats include PostScript, PDF, proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

2. VERBATIM COPYING

   You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

   You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

   If you publish printed copies of the Document numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

   If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a publicly-accessible computer-network location containing a complete Transparent copy of the Document, free of added material, which the general network-using public has access to download anonymously at no charge using public-standard network protocols. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has less than five).
C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
D. Preserve all the copyright notices of the Document.
E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
H. Include an unaltered copy of this License.
I. Preserve the section entitled "History", and its title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
K. In any section entitled "Acknowledgements" or "Dedications", preserve the section's title, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
M. Delete any section entitled "Endorsements". Such a section may not be included in the

Modified Version.
N. Do not retitle any existing section as "Endorsements" or to conflict in title with any Invariant Section.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

   You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice.

   The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

   In the combination, you must combine any sections entitled "History" in the various original documents, forming one section entitled "History"; likewise combine any sections entitled "Acknowledgements", and any sections entitled "Dedications". You must delete all sections entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

   You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

   You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

   A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, does not as a whole count as a Modified Version of the Document, provided no compilation copyright is claimed for the compilation. Such a compilation is called an "aggregate", and this License does not apply to the other self-contained works thus compiled with the Document, on account of their being thus compiled, if they are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one quarter of the entire aggregate, the Document's Cover Texts may be placed on covers that surround only the Document within the aggregate. Otherwise they must appear on covers around the whole aggregate.

8. TRANSLATION

   Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License provided that you also include the original English version of this License. In case of a disagreement between the translation and the original English version of this License, the original English version will prevail.

9. TERMINATION

   You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

    The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

    Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  year  your name.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.1
or any later version published by the Free Software Foundation;
with the Invariant Sections being list their titles, with the
Front-Cover Texts being list, and with the Back-Cover Texts being list.
A copy of the license is included in the section entitled ``GNU
Free Documentation License''.
```

If you have no Invariant Sections, write “with no Invariant Sections” instead of saying which ones are invariant. If you have no Front-Cover Texts, write “no Front-Cover Texts” instead of “Front-Cover Texts being *list*”; likewise for Back-Cover Texts.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Concept Index

N

O

P

Q

R

S

T

U

V

W

Table of Contents