GNU Wget

The noninteractive downloading utility
Updated for Wget 1.5.3, Sep 1998

by Hrvoje Nikšić

1 Overview

GNU Wget is a freely available network utility to retrieve files from the World Wide Web, using HTTP (Hyper Text Transfer Protocol) and FTP (File Transfer Protocol), the two most widely used Internet protocols. It has many useful features to make downloading easier, some of them being:

- 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 is capable of descending recursively through the structure of HTML documents and FTP directory trees, making a local copy of the directory hierarchy similar to the one on the remote server. This feature can be used to mirror archives and home pages, or traverse the web in search of data, like a WWW robot (See Section 9.1 [Robots], page 33). In that spirit, Wget understands the `norobots` convention.
- 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 works exceedingly well on slow or unstable connections, retrying the document until it is fully retrieved, or until a user-specified retry count is surpassed. It will try to resume the download from the point of interruption, using `REST` with FTP and `Range` with HTTP servers that support them.
- By default, 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 13).
- 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 20). 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 [Copying], page 37).

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 20), 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 wide-spreadedness.

[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.cc.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 20).

```
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 20). 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 31 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.

2.5 Download Options

'`-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`'
: Do not clobber existing files when saving to directory hierarchy within recursive retrieval of several files. This option is *extremely* useful when you wish to continue where you left off with retrieval of many files. If the files have the '`.html`' or (yuck) '`.htm`' suffix, they will be loaded from the local disk, and parsed as if they have been retrieved from the Web.

'`-c`'
'`--continue`'
: Continue getting an existing file. This is useful when you want to finish up the download started by another program, or a previous instance of Wget. Thus you can write:

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

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

Note that you need not specify this option if all you want is Wget to continue retrieving where it left off when the connection is lost—Wget does this by default. You need this option only when you want to continue retrieval of a file already halfway retrieved, saved by another FTP client, or left by Wget being killed.

Without '`-c`', the previous example would just begin to download the remote file to '`ls-lR.Z.1`'. The '`-c`' option is also applicable for HTTP servers that support the `Range` header.

'`--dot-style=`*style*'
: Set the retrieval style to *style*. Wget traces the retrieval of each document by printing dots on the screen, each dot representing a fixed amount of retrieved data. Any number

of dots may be separated in a *cluster*, to make counting easier. This option allows you to choose one of the pre-defined styles, determining the number of bytes represented by a dot, the number of dots in a cluster, and the number of dots on the line.

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). The `micro` style is exactly the reverse; it is suitable for downloading small files, with 128-byte dots, 8 dots per cluster, and 48 dots (6K) per line.

'`-N`'
'`--timestamping`'
: Turn on time-stamping. See Chapter 5 [Time-Stamping], page 17 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.

'`-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.cc.fer.hr/robots.txt`' will save the downloaded file to '`fly.cc.fer.hr/robots.txt`'.

'`-nH`'
'`--no-host-directories`'
: Disable generation of host-prefixed directories. By default, invoking Wget with '`-r http://fly.cc.fer.hr/`' will create a structure of directories beginning with '`fly.cc.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

'--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 35.

'-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.

'--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.cc.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.

'-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.

NOTE that Netscape Communications Corp. has claimed that false transmissions of '`Mozilla`' as the `User-Agent` are a copyright infringement, which will be prosecuted. **DO NOT** misrepresent Wget as Mozilla.

2.8 FTP Options

'`--retr-symlinks`'
: Retrieve symbolic links on FTP sites as if they were plain files, i.e. don't just create links locally.

'`-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.cc.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.

2.9 Recursive Retrieval Options

'`-r`'
'`--recursive`'
: Turn on recursive retrieving. See Chapter 3 [Recursive Retrieval], page 12 for more details.

'`-l` *depth*'
'`--level=`*depth*'
: Specify recursion maximum depth level *depth* (See Chapter 3 [Recursive Retrieval], page 12). 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 proxy, e.g.:

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

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

'-k'
'--convert-links'
Convert the non-relative links to relative ones locally. Only the references to the documents actually downloaded will be converted; the rest will be left unchanged.

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

'-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'.

'-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 to access the full remote file list when running a mirror, or for debugging purposes.

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 4.5 [Types of Files], page 14 for more details).

'-D *domain-list*'
'--domains=*domain-list*'
Set domains to be accepted and DNS looked-up, where *domain-list* is a comma-separated list. Note that it does *not* turn on '-H'. This option speeds things up, even if only one host is spanned (See Section 4.3 [Domain Acceptance], page 14).

'--exclude-domains *domain-list*'
Exclude the domains given in a comma-separated *domain-list* from DNS-lookup (See Section 4.3 [Domain Acceptance], page 14).

'-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 4.1 [Relative Links], page 13).

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

'-H'
'--span-hosts'
Enable spanning across hosts when doing recursive retrieving (See Section 4.4 [All Hosts], page 14).

'-I *list*'

'--include-directories=*list*'

Specify a comma-separated list of directories you wish to follow when downloading (See Section 4.6 [Directory-Based Limits], page 15 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 4.6 [Directory-Based Limits], page 15 for more details.) Elements of *list* may contain wildcards.

'-nh'

'--no-host-lookup'

Disable the time-consuming DNS lookup of almost all hosts (See Section 4.2 [Host Checking], page 13).

'-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 4.6 [Directory-Based Limits], page 15 for more details.

3 Recursive Retrieval

GNU Wget is capable of traversing parts of the Web (or a single HTTP or FTP server), depth-first following links and directory structure. This is called *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.

The maximum *depth* to which the retrieval may descend is specified with the '`-l`' option (the default maximum depth is five layers). See Chapter 3 [Recursive Retrieval], page 12.

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.

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 invoking recursion may cause grave overloading on your system, because of the fast exchange of data through the network; all of this may hamper other users' work. The same stands for the foreign server you are mirroring—the more requests it gets in a rows, the greater is its load.

Careless retrieving can also fill your file system unctrollably, which can grind the machine to a halt.

The load can be minimized by lowering the maximum recursion level ('`-l`') and/or by lowering the number of retries ('`-t`'). You may also consider using the '`-w`' option to slow down your requests to the remote servers, as well as the numerous options to narrow the number of followed links (See Chapter 4 [Following Links], page 13).

Recursive retrieval is a good thing when used properly. Please take all precautions not to wreak havoc through carelessness.

4 Following Links

When retrieving recursively, one does not wish to retrieve the 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.cc.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 Relative Links

When only relative links are followed (option '`-L`'), recursive retrieving will never span hosts. No time-expensive DNS-lookups will be performed, and the process will be very fast, with the minimum strain of the network. This will suit your needs often, especially when mirroring the output of various `x2html` converters, since they generally output relative links.

4.2 Host Checking

The drawback of following the relative links solely is that humans often tend to mix them with absolute links to the very same host, and the very same page. In this mode (which is the default mode for following links) all URLs the that refer to the same host will be retrieved.

The problem with this option are the aliases of the hosts and domains. Thus there is no way for Wget to know that '`regoc.srce.hr`' and '`www.srce.hr`' are the same host, or that '`fly.cc.fer.hr`' is the same as '`fly.cc.etf.hr`'. Whenever an absolute link is encountered, the host is DNS-looked-up with `gethostbyname` to check whether we are maybe dealing with the same hosts. Although the results of `gethostbyname` are cached, it is still a great slowdown, e.g. when dealing with large indices of home pages on different hosts (because each of the hosts must be and DNS-resolved to see whether it just *might* an alias of the starting host).

To avoid the overhead you may use '`-nh`', which will turn off DNS-resolving and make Wget compare hosts literally. This will make things run much faster, but also much less reliable (e.g. '`www.srce.hr`' and '`regoc.srce.hr`' will be flagged as different hosts).

Note that modern HTTP servers allows one IP address to host several *virtual servers*, each having its own directory hieratchy. Such "servers" are distinguished by their hostnames (all of which point to the same IP address); for this to work, a client must send a `Host` header, which is what Wget does. However, in that case Wget *must not* try to divine a host's "real" address, nor try to use the same hostname for each access, i.e. '`-nh`' must be turned on.

In other words, the '`-nh`' option must be used to enabling the retrieval from virtual servers distinguished by their hostnames. As the number of such server setups grow, the behavior of '`-nh`' may become the default in the future.

4.3 Domain Acceptance

With the '-D' option you may specify the domains that will be followed. The hosts the domain of which is not in this list will not be DNS-resolved. Thus you can specify '-Dmit.edu' just to make sure that **nothing outside of MIT gets looked up**. This is very important and useful. It also means that '-D' does *not* imply '-H' (span all hosts), which must be specified explicitly. Feel free to use this options since it will speed things up, with almost all the reliability of checking for all hosts. Thus you could invoke

```
wget -r -D.hr http://fly.cc.fer.hr/
```

to make sure that only the hosts in '.hr' domain get DNS-looked-up for being equal to 'fly.cc.fer.hr'. So 'fly.cc.etf.hr' will be checked (only once!) and found equal, but 'www.gnu.ai.mit.edu' will not even be checked.

Of course, domain acceptance can be used to limit the retrieval to particular domains with spanning of hosts in them, but then you must specify '-H' explicitly. E.g.:

```
wget -r -H -Dmit.edu,stanford.edu http://www.mit.edu/
```

will start with 'http://www.mit.edu/', following links across MIT and Stanford.

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/
```

4.4 All Hosts

When '-H' is specified without '-D', all hosts are freely spanned. There are no restrictions whatsoever as to what part of the net Wget will go to fetch documents, other than maximum retrieval depth. If a page references 'www.yahoo.com', so be it. Such an option is rarely useful for itself.

4.5 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.

4.6 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 *upper* 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.

4.7 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 remote and local files. Such information are called the *time-stamps*.

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`'.

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 is newer, the remote file will not be re-fetched. However, if the remote file is more recent, Wget will proceed fetching it normally.

The same goes for FTP. For example:

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

`ls` will show that the timestamps are set according to the state on the remote server. Reissuing the command with '`-N`' will make Wget re-fetch *only* the files that have been modified.

In both HTTP and FTP retrieval Wget will time-stamp the local file correctly (with or without '`-N`') if it gets the stamps, i.e. gets the directory listing for FTP or the `Last-Modified` header for HTTP.

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

```
wget --timestamping -r ftp://prep.ai.mit.edu/pub/gnu/
```

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]

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 received from the directory listings.

For each directory files must be retrieved from, Wget will use the `LIST` command to get the listing. It will try to analyze the listing, assuming that it is a Unix `ls -l` listing, and extract the time-stamps. The rest is exactly the same as for HTTP.

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.

[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.

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, the letter after '=' denoting the value the command takes. It is 'on/off' for 'on' or 'off' (which can also be '1' or '0'), *string* for any non-empty string or *n* for a positive integer. For example, you may specify 'use_proxy = off' to disable use of proxy servers by default. You may use 'inf' for infinite values, where appropriate.

Most of the commands have their equivalent command-line option (See Chapter 2 [Invoking], page 2), except some more obscure or rarely used ones.

accept/reject = *string*
: Same as '-A'/'-R' (See Section 4.5 [Types of Files], page 14).

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

continue = on/off
: Enable/disable continuation of the retrieval, the same as '-c' (which enables it).

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

base = *string*
: Set base for relative URLs, the same as '-B'.

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'.

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 4.3 [Domain Acceptance], page 14).

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).

dot_style = *string*
: Specify the dot retrieval *style*, as with '`--dot-style`'.

exclude_directories = *string*
: Specify a comma-separated list of directories you wish to exclude from download, the same as '`-X`' (See Section 4.6 [Directory-Based Limits], page 15).

exclude_domains = *string*
: Same as '`--exclude-domains`' (See Section 4.3 [Domain Acceptance], page 14).

follow_ftp = on/off
: Follow FTP links from HTML documents, the same as '`-f`'.

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`'.

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`'.

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 4.6 [Directory-Based Limits], page 15).

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'.

passive_ftp = on/off
: Set passive FTP, the same as '--passive-ftp'.

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

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'.

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

quota = *quota*
: Specify the download quota, which is useful to put in 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 4.1 [Relative Links], page 13).

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 33). 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'.

simple_host_check = on/off
: Same as '-nh' (See Section 4.2 [Host Checking], page 13).

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 17).

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'.

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 all the lines are commented out. For any line to have effect, you must remove the '#' prefix at the beginning of 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 any of the settings in this file, you will have to uncomment
## them (and probably change them).


##
## 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


##
## 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 proxy for Wget to use.  It will override the
# value in the environment.
#http_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 have Wget follow FTP links from HTML files by default, set this
# to on:
#follow_ftp = off
```

7 Examples

The examples are classified into three sections, because of clarity. The first section is a tutorial for beginners. The second section explains some of the more complex program features. The third section contains advice for mirror administrators, as well as even more complex features (that some would call perverted).

7.1 Simple Usage

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

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

  The response will be something like:

  ```
  --13:30:45--  http://fly.cc.fer.hr:80/en/
             => 'index.html'
  Connecting to fly.cc.fer.hr:80... connected!
  HTTP request sent, awaiting response... 200 OK
  Length: 4,694 [text/html]

      OK -> ....                                                   [100%]

  13:30:46 (23.75 KB/s) - 'index.html' saved [4694/4694]
  ```

- 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.cc.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.cc.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.cc.fer.hr/welcome.msg
  --10:08:47--  ftp://gnjilux.cc.fer.hr:21/welcome.msg
             => 'welcome.msg'
  Connecting to gnjilux.cc.fer.hr:21... connected!
  Logging in as anonymous ... Logged in!
  ==> TYPE I ... done.  ==> CWD not needed.
  ==> PORT ... done.    ==> RETR welcome.msg ... done.
  Length: 1,340 (unauthoritative)

      OK -> .                                                      [100%]

  10:08:48 (1.28 MB/s) - 'welcome.msg' saved [1340]
  ```

- 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/
lynx index.html
```

7.2 Advanced Usage

- You would like to read the list of URLs from a file? Not a problem with that:

  ```
  wget -i file
  ```

  If you specify '-' as file name, the URLs will be read from standard input.
- Create a mirror image of GNU WWW 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 -t1 http://www.gnu.ai.mit.edu/ -o gnulog
  ```
- Retrieve the first layer of yahoo links:

  ```
  wget -r -l1 http://www.yahoo.com/
  ```
- 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:

  ```
  wget -s http://www.lycos.com/
  more index.html
  ```
- Retrieve the first two levels of '`wuarchive.wustl.edu`', saving them to /tmp.

  ```
  wget -P/tmp -l2 ftp://wuarchive.wustl.edu/
  ```
- You want to download all the GIFs from an HTTP directory. '`wget http://host/dir/*.gif`' doesn't work, since HTTP retrieval does not support globbing. In that case, use:

  ```
  wget -r -l1 --no-parent -A.gif http://host/dir/
  ```

  It is a bit of a kludge, but it works. '`-r -l1`' means to retrieve recursively (See Chapter 3 [Recursive Retrieval], page 12), with maximum depth of 1. '`--no-parent`' means that references to the parent directory are ignored (See Section 4.6 [Directory-Based Limits], page 15), 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.ai.mit.edu/
  ```
- 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@jagor.srce.hr/.emacs
  ```
- If you do not like the default retrieval visualization (1K dots with 10 dots per cluster and 50 dots per line), you can customize it through dot settings (See Section 6.3 [Wgetrc Commands], page 21). For example, many people like the "binary" style of retrieval, with 8K dots and 512K lines:

  ```
  wget --dot-style=binary ftp://prep.ai.mit.edu/pub/gnu/README
  ```

  You can experiment with other styles, like:

  ```
  wget --dot-style=mega ftp://ftp.xemacs.org/pub/xemacs/xemacs-20.4/xemacs-20.4.tar.gz
  wget --dot-style=micro http://fly.cc.fer.hr/
  ```

  To make these settings permanent, put them in your '`.wgetrc`', as described before (See Section 6.4 [Sample Wgetrc], page 24).

7.3 Guru Usage

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

  ```
  crontab
  0 0 * * 0 wget --mirror ftp://ftp.xemacs.org/pub/xemacs/ -o /home/me/weeklog
  ```

- You may wish to do the same with someone's home page. But you do not want to download all those images—you're only interested in HTML.

  ```
  wget --mirror -A.html http://www.w3.org/
  ```

- But what about mirroring the hosts networkologically close to you? It seems so awfully slow because of all that DNS resolving. Just use '-D' (See Section 4.3 [Domain Acceptance], page 14).

  ```
  wget -rN -Dsrce.hr http://www.srce.hr/
  ```

  Now Wget will correctly find out that 'regoc.srce.hr' is the same as 'www.srce.hr', but will not even take into consideration the link to 'www.mit.edu'.

- You have a presentation and would like the dumb absolute links to be converted to relative? Use '-k':

  wget -k -r *URL*

- You would like the output documents to go to standard output instead of to files? OK, but Wget will automatically shut up (turn on '--quiet') to prevent mixing of Wget output and the retrieved documents.

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

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

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

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 HTTP 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.srce.hr`' 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 1.5.3 can be found at '`ftp://prep.ai.mit.edu/pub/gnu/wget-1.5.3.tar.gz`'

8.3 Mailing List

Wget has its own mailing list at ⟨`wget@sunsite.auc.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.auc.dk`⟩. the magic word '`subscribe`' in the subject line. Unsubscribe by mailing to ⟨`wget-unsubscribe@sunsite.auc.dk`⟩.

The mailing list is archived at '`http://fly.cc.fer.hr/archive/wget`'.

8.4 Reporting Bugs

You are welcome to send bug reports about GNU Wget to ⟨`bug-wget@gnu.org`⟩. The bugs that you think are of the interest to the public (i.e. more people should be informed about them) can be Cc-ed to the mailing list at ⟨`wget@sunsite.auc.dk`⟩.

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 on '`wget -rLl0 -t5 -Y0 http://yoyodyne.com -o /tmp/log`', you should try to see if it will crash with a simpler set of options.

   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, should you 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.

5. Find where the bug is, fix it and send me the patches. :-)

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.auc.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, like the Robots Exclusion Standard specification, as well as a list of contributors to GNU Wget.

9.1 Robots

Since Wget is able to traverse the web, it counts as one of the Web *robots*. Thus Wget understands *Robots Exclusion Standard* (RES)—contents of '/robots.txt', used by server administrators to shield parts of their systems from wanderings of Wget.

Norobots support is turned on only when retrieving recursively, and *never* for the first page. Thus, you may issue:

```
wget -r http://fly.cc.fer.hr/
```

First the index of fly.cc.fer.hr will be downloaded. If Wget finds anything worth downloading on the same host, only *then* will it load the robots, and decide whether or not to load the links after all. '/robots.txt' is loaded only once per host. Wget does not support the robots META tag.

The description of the norobots standard was written, and is maintained by Martijn Koster ⟨m.koster@webcrawler.com⟩. With his permission, I contribute a (slightly modified) texified version of the RES.

9.1.1 Introduction to RES

WWW Robots (also called *wanderers* or *spiders*) are programs that traverse many pages in the World Wide Web by recursively retrieving linked pages. For more information see the robots page.

In 1993 and 1994 there have been occasions where robots have visited WWW servers where they weren't welcome for various reasons. Sometimes these reasons were robot specific, e.g. certain robots swamped servers with rapid-fire requests, or retrieved the same files repeatedly. In other situations robots traversed parts of WWW servers that weren't suitable, e.g. very deep virtual trees, duplicated information, temporary information, or cgi-scripts with side-effects (such as voting).

These incidents indicated the need for established mechanisms for WWW servers to indicate to robots which parts of their server should not be accessed. This standard addresses this need with an operational solution.

This document represents a consensus on 30 June 1994 on the robots mailing list (robots@webcrawler.com), between the majority of robot authors and other people with an interest in robots. It has also been open for discussion on the Technical World Wide Web mailing list (www-talk@info.cern.ch). This document is based on a previous working draft under the same title.

It is not an official standard backed by a standards body, or owned by any commercial organization. It is not enforced by anybody, and there no guarantee that all current and future robots will use it. Consider it a common facility the majority of robot authors offer the WWW community to protect WWW server against unwanted accesses by their robots.

The latest version of this document can be found at 'http://info.webcrawler.com/mak/projects/robots/norobots.html'.

9.1.2 RES Format

The format and semantics of the '/robots.txt' file are as follows:

The file consists of one or more records separated by one or more blank lines (terminated by CR, CR/NL, or NL). Each record contains lines of the form:

```
<field>:<optionalspace><value><optionalspace>
```

The field name is case insensitive.

Comments can be included in file using UNIX bourne shell conventions: the '#' character is used to indicate that preceding space (if any) and the remainder of the line up to the line termination is discarded. Lines containing only a comment are discarded completely, and therefore do not indicate a record boundary.

The record starts with one or more User-agent lines, followed by one or more Disallow lines, as detailed below. Unrecognized headers are ignored.

The presence of an empty '/robots.txt' file has no explicit associated semantics, it will be treated as if it was not present, i.e. all robots will consider themselves welcome.

9.1.3 User-Agent Field

The value of this field is the name of the robot the record is describing access policy for.

If more than one User-agent field is present the record describes an identical access policy for more than one robot. At least one field needs to be present per record.

The robot should be liberal in interpreting this field. A case insensitive substring match of the name without version information is recommended.

If the value is '*', the record describes the default access policy for any robot that has not matched any of the other records. It is not allowed to have multiple such records in the '/robots.txt' file.

9.1.4 Disallow Field

The value of this field specifies a partial URL that is not to be visited. This can be a full path, or a partial path; any URL that starts with this value will not be retrieved. For example, 'Disallow: /help' disallows both '/help.html' and '/help/index.html', whereas 'Disallow: /help/' would disallow '/help/index.html' but allow '/help.html'.

Any empty value, indicates that all URLs can be retrieved. At least one Disallow field needs to be present in a record.

9.1.5 Norobots Examples

The following example '`/robots.txt`' file specifies that no robots should visit any URL starting with '`/cyberworld/map/`' or '`/tmp/`':

```
# robots.txt for http://www.site.com/

User-agent: *
Disallow: /cyberworld/map/ # This is an infinite virtual URL space
Disallow: /tmp/ # these will soon disappear
```

This example '`/robots.txt`' file specifies that no robots should visit any URL starting with '`/cyberworld/map/`', except the robot called '`cybermapper`':

```
# robots.txt for http://www.site.com/

User-agent: *
Disallow: /cyberworld/map/ # This is an infinite virtual URL space

# Cybermapper knows where to go.
User-agent: cybermapper
Disallow:
```

This example indicates that no robots should visit this site further:

```
# go away
User-agent: *
Disallow: /
```

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@srce.hr`⟩. 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 the mailing list and the initial FTP space.
- Shawn McHorse—bug reports and patches.
- Kaveh R. Ghazi—on-the-fly `ansi2knr`-ization.
- 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, plust 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.
- Brian Gough—a generous donation.

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:

Tim Adam, Martin Baehr, Dieter Baron, Roger Beeman and the Gurus at Cisco, Mark Boyns, John Burden, Wanderlei Cavassin, Gilles Cedoc, Tim Charron, Noel Cragg, Kristijan Čonkaš, Damir Džeko, Andrew Davison, Ulrich Drepper, Marc Duponcheel, Aleksandar Erkalović, Andy Eskilsson, Masashi Fujita, Howard Gayle, Marcel Gerrits, Hans Grobler, Mathieu Guillaume, Karl Heuer, Gregor Hoffleit, Erik Magnus Hulthen, Richard Huveneers, Simon Josefsson, Mario Jurić, Goran Kezunović, Robert Kleine, Fila Kolodny, Alexander Kourakos, Martin Kraemer, *Σίμος Ξενιτέλλης* (Simos KSenitellis), Tage Stabell-Kulo, Hrvoje Lacko, Dave Love, Jordan Mendelson, Lin Zhe Min, Charlie Negyesi, Andrew Pollock, Steve Pothier, Marin Purgar, Jan Prikryl, Keith Refson, Tobias Ringstrom, Juan José Rodrígues, Heinz Salzmann, Robert Schmidt, Toomas Soome, Sven Sternberger, Markus Strasser, Szakacsits Szabolcs, Mike Thomas, Russell Vincent, 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.

GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

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.

Concept Index

Table of Contents