

Интерактивная система просмотра системных руководств (man-ов)

Тема

[\[Список руководств | Печать\]](#)

Набор

Категория

wget (1)

- >> [wget](#) (1) (Русские man: Команды и прикладные программы пользовательского уровня)
- [wget](#) (1) (Linux man: Команды и прикладные программы пользовательского уровня)
- Ключ [wget](#) обнаружен в базе ключевых слов.

НАЗВАНИЕ

wget - руководство GNU Wget

СИНТАКСИС

wget [*параметры*]... [*URL*]...

ОПИСАНИЕ

GNU Wget - это открыто распространяемая утилита для загрузки файлов из интернет. Она поддерживает протоколы HTTP, HTTPS, и FTP, загрузку с серверов прокси по протоколу HTTP.

Wget может следовать по ссылкам страниц HTML и создавать локальные копии удаленных сайтов web, при этом возможно полное восстановление структуры папок сайта ("recursive downloading" - рекурсивная загрузка). Во время такой работы Wget ищет файл с правами доступа для роботов (/robots.txt). Возможна также конвертация ссылок в загруженных файлах HTML для дальнейшего просмотра сайта в автономном режиме ("off-line browsing").

Проверка заголовков файлов: Wget может считывать заголовки файлов (это доступно по протоколам HTTP и FTP) и сравнивать их с заголовкам ранее загруженных файлов, после чего может загрузить новые версии файлов. Благодаря этому при использовании Wget можно реализовывать зеркальное хранение сайтов или набора файлов на FTP.

Wget разработан для медленных или нестабильных соединений: если во время загрузки возникнет проблема, то Wget будет пытаться продолжить загрузку файла. Если сервер, с

которого загружается файл, поддерживает докачку файлоа, то Wget продолжит загружать файл именно с того места, где оборвалась загрузка.

ОПЦИИ

Основные параметры

-V

--version

Отобразить версию Wget.

-h

--help

Отобразить параметры командной строки Wget.

-b

--background

Перейти в фоновый режим после запуска. Если файл для сообщений не указан параметром **-o**, то он записывается в *wget-log*

-e *command*

--execute *command*

Выполнить *command* как если бы она была частью *.wgetrc*. Команда будет выполнена после команд в *.wgetrc*.

Параметры сообщений

-o *logfile*

--output-file=*logfile*

Записывать все сообщения в *logfile*. Иначе они будут направлены в stderr.

-a *logfile*

--append-output=*logfile*

Дополнить *logfile*. Как и **-o**, только *logfile* не заменяется, а дополняется. Если *logfile* не существует, создается новый файл.

-d

--debug

Отображать сообщения отладки - разная информация, важная для разработчиков Wget.

-q

--quiet

Выключить сообщения Wget.

-v

--verbose

Включить подробные сообщения, со всеми доступными данными. По умолчанию включено.

-nv

--non-verbose

Использовать сокращенные сообщения (чтобы выключить сообщения, см. **-q**). Сообщения об ошибках и основная информация будут отображаться.

-i file

--input-file=file

Считать URL из *file*. В этом случае указывать URL в командной строке не требуется. Если URL указаны и в командной строке и в *file*, то сначала будут загружены URL из командной строки. *file* не обязательно должен иметь формат HTML (но ничего страшного, если это так) -- URL просто должны быть указаны в нем.

Если вы укажете **--force-html**, то файл будет прочтен как **html**. В этом случае могут возникнуть проблемы с относительными ссылками. Это можно предотвратить добавлением в него "<base href='url'>" или вводом в командной строке **--base=url**.

-F

--force-html

При чтении URL из файла, включает чтение файла как HTML. Для предотвращения ошибок в случае локального файла HTML добавьте в файл "<base href='url'>" или введите параметр командной строки **--base**.

-B URL

--base=URL

При чтении URL из файла (**-F**) определяет *URL*, добавляемый к относительным адресам файла, указанного параметром **-i**.

Параметры загрузки

--bind-address=ADDRESS

При соединениях по протоколам TCP/IP передает "bind()" в *ADDRESS* на локальной машине. В *ADDRESS* может быть указано как имя хоста, так и IP адрес. Используется, если ваш компьютер имеет несколько адресов IP.

-t number**--tries=number**

Устанавливает число повторов *number*. Укажите 0 или *inf* для отмены повторов.

-O file**--output-document=file**

Документы не будут записываться в соответствующие файлы, а будут объединены вместе и записаны в файл *file*. Если *file* существует, то он будет заменен. Если *file* указан, как -, то документы будут выводиться на стандартный вывод (stdout). Этот параметр автоматически устанавливает число повторов в 1. Полезен при загрузке разделенных на части файлов с почтовых серверов через интерфейс web.

-nc**--no-clobber**

Если при загрузке сайта связь оборвалась, то укажите этот параметр для продолжения загрузки с места прерывания связи.

При запуске Wget без параметров **-N**, **-nc**, или **-r** загрузка одного и того же файла в одну папку приведет к созданию копии файла с именем *file.1*. Если существует файл с таким именем, третья копия будет называться *file.2* и т.д. При параметре **-nc** будут выводиться предупреждения об этом.

При запуске Wget с параметром **-r**, но без **-N** или **-nc**, новая загрузка сайта приведет к замене уже загруженных файлов. При указании параметра **-nc** загрузка продолжится с места обрыва и загруженные файлы не будут загружаться заново (если только они не изменились).

При запуске Wget с параметром **-N**, с или без **-r**, файл будет загружен только если он новее уже существующего, или если размер его не совпадает с имеющейся копией (см. Сравнение по дате). **-nc** не комбинируется с **-N**.

При указанном параметре **-nc** файлы с расширениями **.html** или (вот это просто ужасно) **.htm** с локальных дисков будут загружаться, как будто бы из интернет.

-c**--continue**

Возобновление загрузки файла. Используется, если загрузка файла была прервана. Например:

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

Если в текущей папке уже есть файл с именем *ls-lR.Z*, то Wget проверит, соответствует ли данный файл загружаемому (не по размеру!), и если это так, то отправит запрос на сервер на продолжение загрузки файла с того же места, где оборвалась загрузка в прошлый раз.

Помните, что при обрыве связи Wget повторяет попытки дозагрузки самостоятельно и без параметра **-с**, и только когда он "сдастся" и завершит свою работу, то для возобновления загрузки файла будет нужен этот параметр.

Без указания опции **-с** предыдущий пример приведет к загрузке указанного файла заново с конечным именем *ls-lR.Z.1*, никак не трогая уже имеющийся *ls-lR.Z*.

Начиная с версии 1.7 при указании параметра **-с** если файл на сервере имеет равный или меньший, чем у локального файла, размер, то Wget не будет ничего загружать и отобразит соответствующее сообщение.

Однако, при использовании **-с** любой файл на сервере, имеющий больший размер, чем локальный файл, будет рассматриваться, как недокачанный. При этом будет загружено и записано в конец файла только "(длина(удал.файл) - длина(локал.файл))" байт. Это может пригодиться, если вам нужно загрузить новые сообщения из какого-либо журнала (log).

При этом если загружаемый файл больше потому, что он *изменился*, то вы получите поврежденный файл (т.е. файл может в итоге получиться совершенно отличным от оригинала). Нужно быть особенно внимательным при использовании **-с** вместе с **-г**, так как каждый измененный файл может быть кандидатом на ``незавершенную загрузку``.

Вы также получите поврежденный файл если ваш сервер прокси HTTP работает глупо и при обрыве соединения пишет в файл сообщение `'transfer interrupted'`. Вероятно, в следующих версиях Wget сам будет исправлять это.

Помните, что **-с** работает только с серверами FTP и HTTP, поддерживающими заголовки "Range" (т.е. докачку файлов).

--progress=type

Индикатор прогресса загрузки и его тип. Возможные значения: ```dot``` и ```bar```.

По умолчанию используется ```bar```. Указание опции **--progress=bar** приведет к прорисовке красивого индикатора из символов ASCII (как `'термометр'`). Если стандартный выход не TTY, то будет использован ```dot```.

Укажите **--progress=dot** чтобы переключится на тип ```dot```. Прогресс загрузки будет отмечаться добавлением в полосу точки или знака равно, каждый символ представляет одинаковое количество данных.

При использовании этого типа, вы можете указать его стиль - **dot:style**. Если стиль "default", то каждый символ будет представлять 1 Кб, 10 символов в кластере и 50 в строке. Стиль "binary" имеет более 'компьютерный' вид - 8Кб в символе, 16 символов в кластере и 48 символов в строке (получается строка на 384 Кб). Стиль "mega" используется для загрузки больших файлов - каждый символ представляет 64Кб, 8 символов в кластере и 48 символов в строке (получается 3 Мб на строку).

Вы можете определить стиль по умолчанию, используя команду "progress" в *.wgetrc*. Если вы хотите, чтобы тип индикатора ```bar``` использовался всегда (а не только при выводе в stdout), то укажите **--progress=bar:force**.

-N

--timestamping

Включить сравнение по дате.

-S

--server-response

Отображать заголовки, отправляемые серверам HTTP и запросы, отправляемые серверам FTP.

--spider

Установка поведения Wget как "паука", т. е. Wget не будет загружать файлы, а только будет проверять их наличие. Так можно проверять закладки и ссылки сайта. Например:

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

Wget не содержит все возможности "настоящих пауков" для WWW.

-T seconds

--timeout=seconds

Время ожидания в секундах. По умолчанию время ожидания равно 900 с (15 мин). Установка значения в 0 отменяет проверку времени ожидания.

Пожалуйста, не снижайте значение времени ожидания, если Вы точно не знаете, что именно делаете.

-W seconds

--wait=seconds

Пауза в секундах между несколькими загрузками (в т.ч. повторами). Это снижает загруженность сервера. Чтобы указать значение в минутах, используйте "m", в часах - "h", в днях - "d" после числа.

Указание большого значения этого параметра полезно, если сеть нестабильна (например при обрывах модемной связи).

--waitretry=seconds

Устанавливает паузу только между повторами оборвавшихся загрузок. Wget будет ждать 1 секунду после первого обрыва, 2 секунды после второго обрыва загрузки того же файла, и т.д. - до максимума, который указывается в секундах. Например, при значении данного параметра, равного 10, Wget будет ждать в общем $(1 + 2 + \dots + 10) = 55$ секунд для каждого файла.

Это значение указывается по умолчанию в файле *wgetrc*.

--random-wait

Некоторые серверы, совершая формирование файлов журналов с паузами запросов файлов, могут определить рекурсивную загрузку файлов - сканирование роботами, такими, как Wget. Этот параметр устанавливает время между запросами, варьируя паузы со временем, рассчитываемым от 0 до $2 * wait$ (секунд), где *wait* указано параметром **-W** для маскировки Wget.

Нельзя забывать, что исходный код Wget доступен, и поэтому даже эту маскировку при желании можно вычислить.

-Y on/off

--proxy=on/off

Поддержка сервера прокси. Включена по умолчанию, если прокси определен.

-Q *quota*

--quota=*quota*

Квота на размер загружаемых файлов. Указывается в байтах (по умолчанию), в килобайтах Кб (если в конце **k**) или в мегабайтах Мб (если в конце **m**).

При исчерпании квоты текущий файл загружается до конца, то есть квота не работает при загрузке одного файла. Например, если вы исполните `wget -Q10k ftp://wuarchive.wustl.edu/ls-lR.gz`, то файл `ls-lR.gz` будет полностью загружен. Также все указанные в командной строке файлы будут обязательно загружены, в отличие от списка файлов в одном файле или как при рекурсивной загрузке.

Указание `0` или `inf` отменит квоту.

Параметры загрузки папок

-nd

--no-directories

Не создавать структуру папок при рекурсивной загрузке. При указанном параметре, все файлы будут загружаться в одну папку. Если файл с данным именем уже существует, то он будет сохранен под именем **ИмяФайла.n**.

-x

--force-directories

Противоположно параметру **-nd** - создавать структуру папок, начиная с главной страницы сервера. Например, `wget -x http://fly.srk.fer.hr/robots.txt` приведет к загрузке файла в папку `fly.srk.fer.hr`.

-nH

--no-host-directories

Не создавать пустые папки в начале структуры. По умолчанию `/pub/xemacs/`. Если вы загрузите ее с параметром **-r**, то она сохранится под именем `ftp.xemacs.org/pub/xemacs/`. С параметром **-nH** из имени начальной папки вырежется `ftp.xemacs.org/`, и она будет называться `pub/xemacs`. А параметр **-cut-dirs** уберет *number* компонентов. Примеры работы параметра **--cut-dirs**:

```
Без параметров      -> ftp.xemacs.org/pub/xemacs/
-nH                  -> pub/xemacs/
-nH --cut-dirs=1     -> xemacs/
-nH --cut-dirs=2     -> .
```

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

Если вы хотите просто избавиться от структуры папок, то вы можете заменить этот параметр на **-nd** и **-P**. В отличие от **-nd**, **-nd** работает с подкаталогами - например, при **-nH --cut-dirs=1** подкаталог *beta/* запишется, как *xemacs/beta*.

-P prefix

--directory-prefix=prefix

Определяет *начальную папку*, в которой будет сохранена структура папок сайта (или просто файлы). По умолчанию этот параметр равен **.** (текущая папка).

Параметры HTTP

-E

--html-extension

Если тип загруженного файла **text/html** и его адрес не оканчивается на **\.[Hh][Tt][Mm][Ll]?**, при использовании данного параметра к его имени будет добавлено **.html**. Это может быть полезно при зеркальном хранении страниц **.asp**, если вы не хотите, чтобы они вмешивались в работу вашего сервера Apache. Другой случай применения этого параметра - это загрузка страниц-ответов CGI скриптов. Страница с URL вида <http://site.com/article.cgi?25> будет сохранена, как *article.cgi?25.html*.

Примечание: при обновлении или другой перезагрузке страниц с данным параметром последние будут загружаться заново в любом случае, т.к. Wget не может узнать, имеет ли отношение локальный файл *X.html* к загружаемому с URL *X*. Чтобы избежать лишней перезагрузки, используйте опции **-k** и **-K**. При этом оригинальные версии файлов будут также сохранены как *X.orig*.

--http-user=user

--http-passwd=password

Имя пользователя *user* и пароль *password* для сервера HTTP. В зависимости от типа отклика, Wget будет использовать "basic" (небезопасную) или "digest" (защищенную) авторизацию.

Можно также указывать имя пользователя и пароль и в самом URL.

-C on/off

--cache=on/off

Включает или выключает кеширование со стороны сервера. При этом Wget посылает соответствующих запрос (**Pragma: no-cache**). Также используется для быстрого обновления файлов на прокси-сервере.

По умолчанию кеширование разрешено.

--cookies=on/off

Включает или выключает использование cookie. Сервер отправляет клиенту cookie, используя заголовок "Set-Cookie" и клиент отвечает таким же cookie. Благодаря этому сервер может вести статистику посетителей. По умолчанию cookie используются, но запись их на диск выключена.

--load-cookies file

Загружать cookie из *file* перед первой загрузкой HTTP. *file* имеет текстовый формат, как *cookies.txt* у Netscape.

Этот параметр используется при зеркалировании. Для этого Wget отправляет те же cookies, которые отправляет ваш браузер при соединении с сервером HTTP. Это включается данным параметром - просто укажите Wget путь к *cookies.txt*. Разные браузеры хранят cookie в разных папках:

Netscape 4.x.

Файл находится в *~/.netscape/cookies.txt*.

Mozilla и Netscape 6.x.

Mozilla хранит cookies в *cookies.txt*, расположенном где-то в *~/.mozilla*, в папке вашего профиля. Полный путь обычно заканчивается чем-то вроде *~/.mozilla/default/some-weird-string/cookies.txt*.

Internet Explorer.

Чтобы экспортировать cookie для Wget, выберите "Файл", "Импорт и Экспорт", в мастере выберите "Экспорт файлов cookie". Проверено в Internet Explorer 5; возможно не будет работать в ранних версиях.

Другие обозреватели.

Параметр **--load-cookies** будет работать с cookie в формате Netscape, который поддерживается Wget.

Если вы не можете использовать параметр **--load-cookies**, то все равно есть выход. Если ваш обозреватель поддерживает Запишите имя и значение cookie и вручную укажите Wget отправку этих cookie:

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

--save-cookies file

Сохранить cookie из *file* в конце сессии. Устаревшие cookie не сохраняются.

--ignore-length

Некоторые серверы HTTP (точнее, скрипты CGI) отправляют заголовки "Content-Length", которые указывают Wget, что загружено еще не все. И Wget загружает один документ несколько раз.

С этим параметром, Wget будет игнорировать заголовки "Content-Length".

--header=*additional-header*

Определяет *additional-header*, отправляемый серверу HTTP. Он должен содержать : и символы после него.

Вы можете определить несколько дополнительных заголовков через использование **--header** несколько раз.

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

Указание пустой строки в значении заголовка очистит все определенные пользователем до этого заголовки.

--proxy-user=*user*

--proxy-passwd=*password*

Определяет имя пользователя *user* и пароль *password* для авторизации сервере прокси. Будет использован тип авторизации "basic".

--referer=*url*

Добавляет заголовок 'Referer: *url*' в запрос HTTP. Используется при загрузке страниц, которые передаются правильно только если сервер знает, с какой страницы вы пришли.

-S

--save-headers

Сохранять заголовки, отправляемые серверам HTTP.

-U *agent-string*

--user-agent=*agent-string*

Идентифицироваться, как *agent-string* при запросе на HTTP сервер.

Протокол HTTP позволяет определять себя использованием заголовка агента. Wget по умолчанию идентифицируется, как **Wget/version**, где *version* - это версия Wget.

Некоторые серверы выдают требуемую информацию только для обозревателей, идентифицирующихся как "Mozilla" или Microsoft "Internet Explorer". Этот параметр позволяет обмануть такие серверы.

Параметры FTP

-nr

--dont-remove-listing

Не удалять временные файлы *.listing*, генерируемые при загрузке по FTP. Эти файлы содержат информацию о папках серверов FTP. Неудаление поможет вам быстро определить обновление папок сервера (т.е. определять, что ваше зеркало является таковым).

Если вы не удаляете *.listing*, то помните о своей безопасности! Например, с таким именем можно создать символическую ссылку на */etc/passwd* или что-то еще.

-g on/off

--glob=on/off

Включает или выключает использование специальных символов (*маски*) по протоколу FTP. Это может быть ***, *?*, *[* и *]*. Например:

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

По умолчанию использование символов маски разрешено, если URL содержит такие символы.

Вы можете также взять URL в кавычки. Это сработает только на серверах Unix FTP (и эмулирующих выход Unix "ls").

--passive-ftp

Включает пассивный режим FTP, когда соединение иницируется клиентом. Используется при наличии firewall.

--retr-symlinks

При рекурсивной загрузке папок FTP файлы, на которые указывают символические ссылки, не загружаются. Данный параметр отключает это.

Параметр **--retr-symlinks** работает сейчас только для файлов, не для папок.

Помните, что этот параметр не работает при загрузке одиночного файла.

Параметры рекурсивной загрузки

-r

--recursive

Включить рекурсивную загрузку.

-l depth

--level=depth

Максимальная глубина рекурсивной загрузки *depth*. По умолчанию ее значение равно 5.

--delete-after

Удалять каждую страницу (локально) *после* ее загрузки. Используется для сохранения новых версий часто запрашиваемых страниц на прокси. Например:

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

Параметр **-r** включает загрузку по умолчанию, параметр **-nd** отключает создание папок.

При указанном параметре **--delete-after** будет игнорироваться параметр **--convert-links**.

-k

--convert-links

После завершения загрузки конвертировать ссылки в документе для просмотра в автономном режиме. Это касается не только видимых ссылок на другие документы, а ссылок на все внешние локальные файлы.

Каждая ссылка изменяется одним из двух способов:

*

Ссылки на файлы, загруженные Wget изменяются на соответствующие относительные ссылки.

Например: если загруженный файл */foo/doc.html*, то ссылка на также загруженный файл */bar/img.gif* будет выглядеть, как *../bar/img.gif*. Этот способ работает, если есть видимое соотношение между папками одного и другого файла.

*

Ссылки на файлы, не загруженные Wget будут изменены на абсолютные адреса этих файлов на удаленном сервере.

Например: если загруженный файл */foo/doc.html* содержит ссылку на */bar/img.gif* (или на *../bar/img.gif*), то ссылка в файле *doc.html* изменится на <http://host/bar/img.gif>.

Благодаря этому, возможен автономный просмотр сайта и файлов: если загружен файл, на который есть ссылка, то ссылка будет указывать на него, если нет - то ссылка будет указывать на его адрес в интернет (если такой существует). При конвертировании используются относительные ссылки, значит вы сможете переносить загруженный сайт в другую папку, не меняя его структуру.

Только после завершения загрузки Wget знает, какие файлы были загружены. Следовательно, при параметре **-k** конвертация произойдет только по завершении загрузки.

-K

--backup-converted

Конвертировать ссылки обратно - убирать расширение **.orig**. Изменяет поведение опции **-N**.

-m

--mirror

Включить параметры для зеркального хранения сайтов. Этот параметр равен нескольким параметрам: **-r -N -l inf -nr**. Для непринудительного хранения зеркальных копий сайтов вы можете использовать данный параметр.

-p

--page-requisites

Загружать все файлы, которые нужны для отображения страниц HTML. Например: рисунки, звук, каскадные стили.

По умолчанию такие файлы не загружаются. Параметры **-r** и **-l**, указанные вместе могут помочь, но т.к. Wget не различает внешние и внутренние документы, то нет гарантии, что загрузится все требуемое.

Например, *1.html* содержит тег "****", со ссылкой на *1.gif*, и тег "**<A>**", ссылающийся на внешний документ *2.html*. Страница *2.html* аналогична, но ее рисунок - *2.gif* и ссылается она на *3.html*. Скажем, это продолжается до определенного числа.

Если будет дана команда:

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

то *1.html*, *1.gif*, *2.html*, *2.gif* и *3.html* загрузятся. Как видим, *3.html* без *3.gif*, т.к. Wget просто считает число прыжков, по которым он перешел, доходит до 2 и останавливается. А при параметрах:

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

Все файлы и рисунок *3.gif* страницы *3.html* загрузятся. Аналогично

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

приведет к загрузке *1.html*, *1.gif*, *2.html* и *2.gif*. Чтобы загрузить одну указанную страницу HTML со всеми ее элементами, просто не указывайте **-r** и **-l**:

```
wget -p http://I<site>/1.html
```

При этом Wget будет себя вести, как при параметре **-r**, но будут загружены страница и ее вспомогательные файлы. Если вы хотите, чтобы вспомогательные файлы на других серверах (т.е.

через абсолютные ссылки) были загружены, используйте:

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

И в завершении, нужно сказать, что для Wget внешняя ссылка - это URL, указанный в тегах "<A>", "<AREA>" и "<LINK>", кроме "<LINK REL='stylesheet'>".

Параметры запрета/разрешения рекурсивной загрузки

-A *acclist* --accept *acclist*

-R *rejlist* --reject *rejlist*

Список имен файлов, разделенных запятыми, которые следует или не следует загружать. Разрешается задание имен файлов по маске.

-D *domain-list*

--domains=*domain-list*

Список доменов *domain-list*, с которых разрешено загружать файлы. Разделяется запятыми. Этот параметр *не* включает -H.

--exclude-domains *domain-list*

Список доменов, с которых *не* разрешено загружать файлы

--follow-ftp

Следовать по ссылкам FTP с HTML страниц. Иначе, ссылки на файлы по протоколу FTP игнорируются.

--follow-tags=*list*

Wget имеет встроенную таблицу тегов HTML, в которых он ищет ссылки на другие файлы. Вы можете указать дополнительные теги в разделенном запятыми списке *list* в этом параметре.

-G *list*

--ignore-tags=*list*

Обратно **--follow-tags**. Для пропуска тегов HTML при рекурсивной загрузке, укажите их в разделенном запятыми списке *list*.

Раньше параметр **-G** был лучшим для загрузки отдельных страниц с их вспомогательными файлами. Вы можете посмотреть, как это было, указав команду

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

Но теперь лучшим параметром для загрузки одной страницы полностью считается **--page-requisites**.

-H

--span-hosts

Разрешает посещать любые сервера, на которые есть ссылка.

-L

--relative

Следовать только по относительным ссылкам. При этом параметре файлы с других серверов точно не будут загружаться.

-I *list*

--include-directories=*list*

Список папок, разделенных запятыми, из которых разрешено загружать файлы. Элементы списка *list* могут содержать символы масок.

-X *list*

--exclude-directories=*list*

Список папок, разделенных запятыми, исключаемых для загрузки (см. Ограничение по папкам). Элементы списка *list* могут содержать символы масок.

-np

--no-parent

Не подниматься выше начального адреса при рекурсивной загрузке.

ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ

Примеры разделены на три категории согласно их сложности.

Просто использование

*

Если вам нужно загрузить URL, то введите:

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

*

Но что же будет, если соединение медленное, а файл длинный? Есть возможность обрыва связи перед завершением загрузки. В этом случае Wget будет продолжать попытки нового соединения, пока не кончится число попыток (по умолчанию 20). Можно изменить это число, например до 45:

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

*

Теперь оставим Wget работать в фоновом режиме, а его сообщения будем записывать в журнал *log*. Долго набирать **--tries**, так что используем **-t**.

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

Символ амперсанда в конце указывает командному интерпретатору продолжать работу, не дожидаясь завершения работы Wget. Чтобы программа делала повторы бесконечно - используйте **-t inf**.

*

Использовать FTP также очень просто. Wget берет на себя все заботы по авторизации.

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

*

Если вы укажете адрес папки, то Wget загрузит листинг этой папки (т.е. файлы и подкаталоги, содержащиеся в ней) и сконвертирует его в формат HTML. Например:

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

Расширенное использование

*

Если у Вас есть файл с URL, которые вы хотите загрузить, то используйте параметр **-i**:

```
wget -i I<file>
```

Если вы укажете **-** вместо имени файла, то URL будут читаться из стандартного ввода (stdin).

*

Создать пятиуровневую копию сайта GNU со структурой папок оригинала, с одной попыткой загрузки, сохранить сообщения в *gnu.log*:


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

*

Как и в примере выше, но с конвертированием ссылки в файлах HTML в локальные, для последующего автономного просмотра:

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

*

Загрузить одну страницу HTML и все файлы, требуемые для отображения последней (напр. рисунки, файлы каскадных стилей и т. д.). Также сконвертировать все ссылки на эти файлы:

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

Страница HTML будет сохранена в www.server.com/dir/page.html и рисунки, каскадные стили и прочее будет сохранено в папке www.server.com/, кроме случая, когда файлы будут загружаться с других серверов.

*

Как и в примере выше, но без папки www.server.com/. Также *все* файлы будут сохранены в подпапках *download/*.

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

*

Загрузить index.html с www.lycos.com, отображая заголовки сервера:

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

*

Сохранить заголовки в файл для дальнейшего использования.

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

*

Загрузить два высших уровня **wuarchive.wustl.edu** в */tmp*.

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

*

Загрузить файлы GIF папки на HTTP сервере. Команда **wget http://www.server.com/dir/*.gif** не будет работать, так как маскировочные символы не поддерживаются при загрузке по протоколу HTTP. Используйте:

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

-r -l1 включает рекурсивную загрузку с максимальной глубиной 1. **--no-parent** выключает следование по ссылкам в родительскую папку, имеющую верхний уровень, **-A.gif** разрешает загружать только файлы с расширением .GIF. **-A '*.gif'** также будет работать.

*

Предположим, что во время рекурсивной загрузки вам нужно было срочно выключить/перезагрузить компьютер. Чтобы не загружать уже имеющиеся файлы, используйте:

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

*

Если вы хотите указать имя пользователя и пароль для сервера HTTP или FTP, используйте соответствующий синтаксис URL:

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

*

Вы хотите, чтобы загружаемые документы шли в стандартный вывод, а не в файлы?

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

Если вы хотите устроить конвейер и загрузить все сайты, ссылки на которые указаны на одной странице:

Профессиональное использование

*

Для хранения зеркала страницы (или папки FTP), то используйте **--mirror (-m)**, что заменяет **-r -l inf -N**. Вы можете добавить Wget в crontab с запросом на проверку обновлений каждое воскресенье:

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

*

Вы также хотите, чтобы ссылки конвертировались в локальные. Но после прочтения этого руководства, вы знаете, что при этом не будет работать сравнение по времени. Укажите Wget оставлять резервные копии HTML файлов перед конвертацией. Команда:

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

*

А если не работает локальный просмотр файлов HTML с расширением, отличным от **.html**, например *index.cgi*, то нужно передать команду на переименование всех таких файлов (content-type = **text/html**) в *имя.html*.

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

С краткими аналогами команд:

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

ФАЙЛЫ

`/usr/local/etc/wgetrc`

По умолчанию это расположение *глобального* файла настроек.

`.wgetrc`

Файл настроек пользователя.

НАЙДЕННЫЕ ОШИБКИ

Вы можете отправлять отчеты об ошибках в GNU Wget на адрес <bug-wget@gnu.org> (на английском).

Перед отправкой:

1.
Убедитесь, что поведение программы действительно ошибочно. Если Wget "вылетает", то это ошибка. Если поведение Wget не соответствует документации, то это ошибка. Если все работает странно, но вы не уверены, как оно должно работать на самом деле, то это тоже может быть ошибкой.
2.
Попытайтесь повторить ситуацию с выдачей ошибки за минимальное количество действий.

Не спешите отправлять `.wgetrc`, попробуйте проделать все действия, приведшие к ошибке с другим файлом настроек (или вообще без него).
3.
Запустите Wget с параметром `-d` и отправьте журнал (или его части). *Намного* легче отыскивать ошибки при наличии таких журналов.
4.
Если Wget выдал ошибку, то попытайтесь запустить его в отладчике, например `"gdb `which wget` core"` и введите `"where"` для получения обратной трассировки.

СМ. ТАКЖЕ

GNU Info для *wget*.

АВТОРЫ

Hrvoje Niksic <hniksic@arsdigita.com>.

ПЕРЕВОД

Н. Шафоростов <admin@program.net.ua>

АВТОРСКИЕ ПРАВА

Copyright (c) 1996, 1997, 1998, 2000, 2001 Free Software Foundation, Inc.

Это свободное программное обеспечение; условия копирования ищите в исходных текстах. НЕ существует никаких гарантий; программа НЕ предназначена для ПРОДАЖ.

Index

[НАЗВАНИЕ](#)

[СИНТАКСИС](#)

[ОПИСАНИЕ](#)

[ОПЦИИ](#)

[Основные параметры](#)

[Параметры сообщений](#)

[Параметры загрузки](#)

[Параметры загрузки папок](#)

[Параметры HTTP](#)

[Параметры FTP](#)

[Параметры рекурсивной загрузки](#)

[Параметры запрета/разрешения рекурсивной загрузки](#)

[ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ](#)

[Просто использование](#)

[Расширенное использование](#)

[Профессиональное использование](#)

[ФАЙЛЫ](#)[НАЙДЕННЫЕ ОШИБКИ](#)[СМ. ТАКЖЕ](#)[АВТОРЫ](#)[ПЕРЕВОД](#)[АВТОРСКИЕ ПРАВА](#)Поиск по тексту MAN-ов:

Партнёры:

При поддержке
inferno solutions*

Хостинг:

**Hoster.ru**
хостинг провайдер[Закладки на сайте](#)
[Проследить за страницей](#)Created 1996-2024 by [Maxim Chirkov](#)
[Добавить](#), [Поддержать](#), [Вебмастеру](#)