

NAME

httpindex – HTTP front-end for SWISH++ indexer

SYNOPSIS

wget [*options*] *URL...* 2>&1 | **httpindex** [*options*]

DESCRIPTION

httpindex is a front-end for **index(1)** to index files copied from remote servers using **wget(1)**. The files (in a copy of the remote directory structure) can be kept, deleted, or replaced with their descriptions after indexing.

OPTIONS**wget Options**

The **wget(1)** options that are **required** are: **-A**, **-nv**, **-r**, and **-x**; the ones that are **highly recommended** are: **-l**, **-nh**, **-t**, and **-w**. (See the EXAMPLE.)

httpindex Options

httpindex accepts the same short options as **index(1)** except for **-H**, **-I**, **-l**, **-r**, **-S**, and **-V**.

The following options are unique to **httpindex**:

- d** Replace the text of local copies of retrieved files with their descriptions after they have been indexed. This is useful to display file descriptions in search results without having to have complete copies of the remote files thus saving filesystem space. (See the `extract_description()` function in **WWW(3)** for details about how descriptions are extracted.)
- D** Delete the local copies of retrieved files after they have been indexed. This prevents your local filesystem from filling up with copies of remote files.

EXAMPLE

To index all HTML and text files on a remote web server keeping descriptions locally:

```
wget -A html,txt -linfo -t2 -rxnv -nh -w2 http://www.foo.com 2>&1 |
httpindex -d -e'html:*.html' -e'text:*.txt'
```

Note that you need to redirect **wget(1)**'s output from standard error to standard output in order to pipe it to **httpindex**.

EXIT STATUS

Exits with a value of zero only if indexing completed successfully; non-zero otherwise.

CAVEATS & FILES

Same as those for **index(1)**.

SEE ALSO

index(1), **wget(1)**, **WWW(3)**

AUTHOR

Paul J. Lucas <pauljlucas@mac.com>