NAME

PyGopherd − Multiprotocol Information Server

SYNOPSIS

pygopherd *[configfile]*

DESCRIPTION

Welcome to **PyGopherd**. In a nutshell, **PyGopherd** is a modern dynamic multi-protocol hierarchical information server with a pluggable modularized extension system, full flexible caching, virtual files and folders, and autodetection of file types -- all with support for standardized yet extensible per-document metadata. Whew! Read on for information on this what all these buzzwords mean.

PyGopherd is designed to serve up files using the Gopher Internet protocol. With Gopher, you can mount a filesystem (viewing files and folders as if they were local), browse Gopherspace with a web browser, download files, and be interactive with searching.

But this is only part of the story. The world of Gopher is more expansive than this. There are two major gopher protocols: Gopher0 (also known as RFC1436) and Gopher+. Gopher0 is a small, simple, lightweight protocol that is very functional yet also extremely easy to implement. Gopher0 clients can be easily places in small embedded devices or in massive environments like a modern web browser.

Gopher+ is based on Gopher0 but extends it by providing document metadata such as file size and MIME type. Gopher+ allows all sorts of neat features, such as configurable metadata (serving up a bunch of photos? Add a Subject field to your metadata to let a customized photo browser display who is pictured) and multiple views of a file (let the user select to view your photos as PNG or JPEG).

PyGopherd supports both.

PyGopherd also fully natively supports HTTP, the protocol used on the Internet for most Web transactions. So, you can access a **PyGopherd** server using anything from a small, 20-line client in mobile phone to a massive 50-MB web browser.

MODULARITY

PyGopherd is designed in an extremely modular fashion. In fact, all protocols in **PyGopherd** and all handlers are modules. You can easily select which modules to use and which to exclude, and the order in which they are tried. Protocol modules implement different ways of speaking to clients (HTTP, Gopher+, Gopher0) and handler modules implement different ways of handling data on the local machine (files, directories, links, mailboxes, etc.)

FEATURE LIST

Here are some of **PyGopherd**'s features:

- Runs on any platform supported by Python 2.2. This includes virtually every past and current flavor of Unix (Linux, *BSD, Solaris, SunOS), Windows, MacOS 9.x and 10.x, and more. Some features may not be available on non-Unix platforms, however.
- Runs on any platform supported by Java 1.1 via the Jython Python implementation.
- Tunable server types via configuration directive -- forking or threading.
- Secure design with support for chrooted execution.
- Feature-complete, full implementations of Gopher0 (RFC1436), Gopher+, and HTTP/1.0.
- Complete support Bucktooth-style gophermap files.
- Complete support for UMN-type .Links, .names, etc. files and support for .cap directories.
- Support for automatically finding the titles of HTML documents for presentation in a directory.

- Virtual folder support can present many different types of files as folders. Some examples are:
  - Can present any Unix MBOX, MMDF box, MH directory, Maildir directory, or Babyl mailbox as a virtual folder, the contents of which are the messages in the mailbox.
  - Can use a configurable separator to split a file into multiple parts, the first line of each becoming the name for the virtual folder.
- Versatile configuration file format is both extensible and nicely complementary of the module system.
- Several options for running external programs to generate dynamic content: UMN-style execution, Bucktooth-style execution, or Pygopher PYG objects.
- Protocol-independant, handler-dependant caching. This speeds up time by letting handlers cache dynamically-generated information -- currently used by the two directory handlers. This can improve performance of directories by several orders of magnitude. Because this is a handler cache only, all protocols can share the single cache -- and since the processing time of the protocols is negligable anyway, it works very well.
- Autosensing of MIME types and gopher0 item types. Both are completely configurable. MIME type detection is done using a standard mime.types file, and gopher0 types are calculated by using a configurable regexp-based MIME-to-gophertype map.
- Heavy support of regular expressions in configuration.
- ProtocolMultiplexer and HandlerMultiplexer let you choose only those protocols and handlers that you wish your server to support and the order in which they are tried when a request comes in.
- Modular architecture permits rapid prototyping of new capabilities.
- Full logging via syslog.

INSTALLATION

If you are reading this document via the "man" command, it is likely that you have no installation tasks to perform; your system administrator has already installed it. If you need to install it yourself, you have three options: a system-wide installation with Debian, system-wide installation with other systems, and a single-user installation. You can download the latest version of **PyGopherd** from http://quux.org/devel/gopher/pygopherd/.

DEBIAN SYSTEM-WIDE INSTALLATION

If you are tracking Debian unstable, you may install **PyGopherD** by simply running the following command as root:

apt-get install pygopherd

If you are not tracking Debian unstable, download the Debian .deb package from the **PyGopherd** website and then run **dpkg -i** to install the downloaded package. Then, go to CONFIGURATION below. You will use **/etc/init.d/pygopherd start** to start the program.

OTHER SYSTEM-WIDE INSTALLATION

Download the tar.gz version of the package from the website. Make sure you have Python 2.2 or above installed; if now, download and install it from www.python.org. Then run these commands:

tar -zxvf pygopherd-x.y.z.tar.gz
cd pygopherd-x.y.z
python2.2 setup.py

Some systems will need to use **python** instead of **python2.2.**

Next, proceed to configuration. Make sure that the */etc/pygopherd/pygopherd.conf* file names valid users (**setuid** and **setgid** options) and a valid document root (**root** option).

You will type **pygopherd** to invoke the program.

SINGLE-ACCOUNT INSTALLATION

Download the tar.gz version of the package from the website. Make sure you have Python 2.2 installed; if not, download it from www.python.org Then run these commands:

tar -zxvf pygopherd-x.y.z.tar.gz
cd pygopherd-x.y.z

Modify *conf/pygopherd.conf* as follows:

- Set **usechroot = no**
- Comment out (add a # sign to the start of the line) the **pidfile**, **setuid**, and **setgid** lines.
- Set **mimetypes = ./conf/mime.types**
- Set **root =** to something appropriate.
- Set **port** to a number greater than 1024.

When you want to run **PyGopherd,** you will issue the **cd** command as above and then type **./bin/pygopherd;** there is no installation step necessary.

CONFIGURATION

PyGopherd is regulated by a configuration file that is normally stored in */etc/pygopherd/pygopherd.conf*. You can specify alternate configuration files on the command-line. The **PyGopherd** distribution ships with a sample *pygopherd.conf* file that thoroughly documents the configuration file options and settings.

OPTIONS

All **PyGopherd** configuration is done via the configuration file. Therefore, the program has only one command-line option.

configfile
This optional argument specifies the location of the configuration file that **PyGopherD** is to use.

HANDLERS

PyGopherd defines several handlers which are responsible for finding data on your server and presenting it to the user. The handlers are used to generate things like links to other documents and directory listings. They are also responsible for serving up regular files and even virtual folders.

Handlers are specified with the **handlers** option in *pygopherd.conf*. This option is a list of handlers to use. For each request that arrives, **PyGopherd** will ask each handler in turn whether or not it can handle the request, and will handle the request according to the first handler that is capable of doing so. If no handlers can handle the request, a file not found error is generated. See the example configuration file for an example.

The remaining parts of this section describe the different handlers that ship with PyGopherd.

dir.DirHandler

This handler is a basic one that is used for directories that contain neither a *gophermap* file nor UMN-style links files, or situations where you have no need for either of those. It simply reads the contents of your on-disk directory, determines the appropriate types of each file, and sends the result to the client. The descriptions of each item are usually set to the filename, but the **html.HTMLFileTitleHandler** may override that.

gophermap.BuckGophermapHandler

This handler is used to generate gopher directory listings based on *gophermap* files. It will not read the directory on-disk, instead serving content from the *gophermap* file only. Gophermaps are useful if you want to present a directory where the files do not frequently change and there is general information to present. Overall, if you only wish to present information particular to certain files, you should consider using the abstract feature of the UMN.UMNDirHandler handler.

The *gophermap* files contain two types of lines, which are described here using the same convention normally used for command-line arguments. In this section, the symbol \t will be used to indicate a tab

character, Control-I.

full line of informational text
gophertypeDESCRIPTION\t[*selector*[\t*host*[\t*port*]]]

The informational text must not contain any tab characters. If present, it will be rendered with gopher type **i**, which willl cause it to be displayed on a client's screen at its particular position in the file.

The second type of line represents a link to a gopher file or directory. It begins with a single-character gopher type (see GOPHER ITEM TYPES below) followed immediately by a description and a tab character. There is no space or other separator between the gopher type and the description.

The remaining arguments are optional, but only to the extent that arguments may be omitted only if all arguments after them are also omitted. These arguments are:

- The *selector* is the name of the file on the gopher server. If it begins with a slash, it is an absolute path; if it does not, it is interpreted relative to the directory that the gophermap file is in. If no selector is specified, the description is also used as the selector.
- The *host* specifies the host on which this resource is located. If not specified, defaults to the current server.
- The *port* specifies the port on which the resource is located. If not specified, defaults to the port the current server is listening on.

An example of a gophermap to help illustrate the concept is included with the **PyGopherd** distribution in the file *examples/gophermap*.

file.CompressedFileHandler

In order to save space, you might want to store documents on-disk in a compressed format. But then clients would ordinarily have to decompress the files themselves. It would be nice to have the server automatically decompress the files on the fly, sending that result to the client. That's where **file.CompressedFileHandler** comes in.

This handler will take compressed files, pipe them through your chosen decompression program, and send the result directly to clients -- completely transparently.

To use this handler, set the **decompressors** option in the configuration file. That option defines a mapping from MIME encodings (as defined with the **encoding** option) to decompression programs. Files that are not encoded, or which have an encoding that does not occur in the **decompressors** map, will not be decompressed by this handler.

Please see the sample configuration file for more examples and details about the configuration of this handler.

file.FileHandler

The **file.FileHandler** is just that -- its duty is to serve up regular files to clients.

html.HTMLFileTitleHandler

This handler is used when generating directories and will set the description of HTML files to the HTML title defined in them rather than let it be the default filename. Other than that, it has no effect. UMN gopherd implements a similar policy.

mbox.MaildirFolderHandler
mbox.MaildirMessageHandler
mbox.MBoxMessageHandler
mbox.MBoxFolderHandler

These four handlers provide a unique "virtual folder" service. They allow you to present mailboxes as if they were folders, the items of the folders being the messages in the mailbox, organized by subject. This is

useful for presenting mail archives or just making e-mail accessible in a nice and easy fashion.

All you have to do to use these handlers is enable them in your **handlers** section. They will automatically detect requests for mailboxes and handle them appropriately.

The different handlers are for traditional Unix mbox mailboxes (all messages in a single file) and new qmail-stype Maildir mailboxes. You can enable only the two handlers for the specific mailbox type that you use, if desired.

pyg.PYGHandler

PYG (short for PYGopherd) is a mechanism that provides a tremendous amount of flexibility. Rather than just letting you execute a script like other Gopher or HTTP servers, PYGs are actually loaded up into PyGopherd and become fully-capable first-class virtual handlers. Yet they need not be known ahead of time, and are loaded dynamically.

With a PYG handler, you can generate gopher directories, handle searches, generate files, and more on the fly. You can create entire virtual directory trees (for instance, to interface with NNTP servers or with DICT servers), and access them all using the standard Gopher protocol. All of this without having to modify even one line of PyGopherd code.

If enabled, the **pyg.PYGHandler** will look for files with the extension .pyg that are marked executable. If found, they will be loaded and run as PYGs.

Please note: this module provides the capability to execute arbitrary code. Please consider the security ramifications of that before enabling it.

See the **virtual.Virtual** handler for more information about passing data to your scripts at runtime.

At present, documentation on writing PYGs is not provided, but you may find examples in the *pygfarm* directory included with the **PyGopherd** distribution.

scriptexec.ExecHandler

This handler implements "old-style" script execution; that is, executing arbitrary programs and piping the result to the client. It is, for the most part, compatible with both scripts written for UMN gopherd and the Bucktooth gopher server. If enabled, it will execute any file that is marked executable in the filesystem. It will normally list scripts as returning plain text, but you may create a custom link to the script that defines it as returning whatever kind of file you desire. Unlike PYGs, this type must be known in advance.

The **scriptexec.ExecHandler** will set environment variables for your scripts to use. They are as follows:

SERVER_NAME
The name of this server as defined in the configuration file or detected from the operating system.

SERVER_PORT
The port this server is listening on.

REMOTE_ADDR
The IP address of the client.

REMOTE_PORT
The port number of the client.

REMOTE_HOST
The same value as **REMOTE_ADDR**.

SELECTOR
The file that was requested; that is, the relative path to this script. If the selector included additional parameters after a ?, they will be included in this string as well.

REQUEST
The "base" part of the selector; that is, the part leading up to the ?.

SEARCHREQUEST
Included only if the client specified search data, this is used if the client is searching for something.

See the **virtual.Virtual** handler for more information about passing data to your scripts at runtime.

Please note: this module provides the capability to execute arbitrary code. Please consider the security ramifications of that before enabling it.

UMN.UMNDirHandler
This is one of the most powerful workhorse handlers in **PyGopherd**. It is designed to emulate most of the ways in which the UMN gopherd distribution generates directories, even going so far as to be bug-compatible in some cases. Generating directories with this handler is often the best general-purpose way to make nice directories in gopherspace.

The remainder of the description of the **UMN.UMNDirHandler**, except for the ABSTRACTS AND INFO section, is lifted directly from the original UMN gopherd documentation, with light editing, because this handler implements it so exactly that there was no point in rewriting all that documentation :-)

LINKS
You can override the default view of a directory as generated by **dir.DirHandler** by creating what are known as *Links* in the Gopher data directory tree.

The ability to make links to other hosts is how gopher distributes itself among multiple hosts. There are two different ways to make a link. The first and simplest is to create a link file that contains the data needed by the server. By default all files in the gopher data directory starting with a period are taken to be link files. A link file can contain multiple links. To define a link you need to put five lines in a link file that define the needed characteristics for the document. Here is an example of a link.

```
Name=Cheese Ball Recipes
Numb=1
Type=1
Port=150
Path=1/Moo/Cheesy
Host=zippy.micro.umn.edu
```

The Name= line is what the user will see when cruising through the database. In this case the name is "Cheese Ball Recipes". The "Type=" defines what kind of document this object is. For a list of all defined types, see GOPHER ITEM TYPES below. For Gopher+ and HTTP, a MIME type is also used, which is determined automatically based on the type you specify.

The "Path=" line contains the selector string that the client will use to retrieve the actual document. The Numb= specifies that this entry should be presented first in the directory list (instead of being alphabetized). The "Numb=" line is optional. If it is present it cannot be the last line of the link. The "Port=" and "Host=" lines specify a fully qualified domain name (FQDN) and a port respectively. You may substitute a plus '+' for these two parameters if you wish. The server will insert the current hostname and the current port when it sees a plus in either of these two fields.

An easy way to retrieve links is to use the Curses Gopher Client. By pressing '=' You can get information suitable for inclusion in a link file.

OVERRIDING DEFAULTS
The server looks for a directory called *.cap* when parsing a directory. The server then checks to see if the

.cap directory contains a file with the same name as the file it's parsing. If this file exists then the server will open it for reading. The server parses this file just like a link file. However instead of making a new object, the parameters inside the .cap/ file are used to override any of the server supplied default values.

For instance say you wanted to change the Title of a text file for gopher, but don't want to change the filename. You also don't want it alphabetized, instead you want it second in the directory listing. You could make a set-aside file in the .cap directory with the same filename that contained the following lines:

```
Name=New Long Cool Name
Numb=2
```

The replacement (and default) for .cap files are extended link files. The equivilant is to create a file that begins with a dot (.) in the *same* directory as the file you wish to override. If the name of the file was *file-to-change* then you could create a file called *.names* with the following contents

```
Path=./file-to-change
Name=New Long Cool Name
Numb=2
```

ADDING COOL LINKS

One cool thing you can do with .Links is to add neato services to your gopher server. Adding a link like this:

```
Name=Cool ftp directory
Type=h
Path=/URL:ftp://hostname/path/
Host=+
Port=+

Name=Cool web site
Type=h
Path=/URL:http://hostname/
Host=+
Port=+
```

Will allow you to link in any FTP or Web site to your gopher.

You can easily add a finger site to your gopher server thusly:

```
Name=Finger information
Type=0
Path=lindner
Host=mudhoney.micro.umn.edu
Port=79
```

HIDING AN ENTRY

This kind of trick may be necessary in some cases, and thus for object "fred", the overriding .names file entry would be:

```
Type=X
Path=./fred
```

by overriding default type to be "X". This kind of hideouts may be usefull, when for some reason there are symlinks (or whatever) in the directory at which **PyGopherd** looks at, and those entries are not desired to

be shown at all.

ABSTRACTS AND INFO
Many modern gopher server maintainers like to intersperse gopher directory listings with other information -- often, additional information about the contents of files in the directory. The gophermap system provides one way to do that, and abstracts used with UMN gopher directories provides another.

Subject to the **abstract_headers** and **abstract_entries** configuration file options, this feature allows you to define that extra information. You can do that by simply creating a file named *filename*.abstract right alongside the regular file in your directory. The file will be interpreted as the abstract. For a directory, create a file named .abstract in the directory. Simple as that!

url.HTMLURLHandler
PyGopherd provides ways for you to link to pages outside Gopherspace -- that is, web pages, FTP sites, and the like. This is accomplished according to the Links to URL specification (see CONFORMING TO below for details). In order to link to a URL (EXCEPT gopher URLs), you create a link of type h (regardless of the actual type of the resource that you are linking to) in your gophermap or .Links file that looks like this:

/URL:http://www.complete.org/

Modern Gopher clients that follow the Links to URL specification will automatically follow that link when you select it. The rest need some help, and that's where this handler comes in.

For Gopher clients that do not follow the Links to URL specification, the **url.HTMLURLHandler** will automatically generate an HTML document for them on the fly. This document includes a refresh code that will send them to the proper page. You should not disable this handler.

url.URLTypeRewriter
Some people wish to serve HTML documents from their Gopher server. One problem with that is that links in Gopherspace include an extra type character at the beginning, whereas links in HTTP do not. This handler will remove the extra type character from HTTP requests that come in, allowing a single relative-to-root link to work for both.

virtual.Virtual
This handler is not intended to ever be used directly, but is used by many other handlers such as the mbox support, PYG handlers, and others. It is used to generate virtual entries in the directory hierarchy -- that is, entries that look normal to a Gopher client, but do not actually correspond to a file on disk.

One special feature of the **virtual.Virtual** handler is that you can send information to it at runtime in a manner similar to a CGI script on the web. You do this by adding a question mark after the regular selector, followed by any arbitrary data that you wish to have sent to the virtual request handler.

GOPHER ITEM TYPES

When you construct links to files via .Links files or gophermap files, or modify the **mapping** in the configuration file, you will need to know these. Items bearing the "not implemented" text are not served up by **PyGopherd** as it ships, generally due to requirements of customized per-site software, but may be served up via PYG extension modules or other gopher servers.

This list was prepared based on RFC1436, the UMN gopherd(1) manpage, and best current practices.

0 Plain text file

1 Directory

2 CSO phone book server (not implemented by **PyGopherd**)

3 Error condition; text that follows in plain text.

4 Macintosh file, BinHex format

5 DOS binary archive (not implemented by **PyGopherd**; use type 9 instead)

6 uuencoded file; not directly generated by **PyGopherd** automatically, but can be linked to manually. Most gopher clients will handle this better as type 1.

7 A Gopher search

8 A telnet link

9 Binary file

+ Redundant server (not implemented by **PyGopherd**)

c Calendar (not implemented by **PyGopherd**)

e Event (not implemented by **PyGopherd**)

g GIF-format graphic

h An HTML file

I Any kind of graphic file other than GIF

i Informational text included in a directory that is displayed but does not link to any actual file.

M A MIME multipart/mixed file

s Any kind of sound file

T tn3270 link

X, − UMN-specific -- signifies that this entry should not be displayed in a directory entry, but may be accessed via a direct link. This value is never transmitted in any Gopher protocol.

CONFORMING TO

- The Internet Gopher Protocol as specified in RFC1436
- The Gopher+ upward-compatible enhancements to the Internet Gopher Protocol from the University of Minnesota as laid out at gopher://gopher.quux.org/0/Archives/mirrors/boombox.micro.umn.edu/pub/gopher/gopher_protocol/Gopher+/Gopher+.txt
- The gophermap file format as originally implemented in the Bucktooth gopher server and described at gopher://gopher.floodgap.com/0/buck/dbrowse%3Ffaquse%201
- The Links to URL specification as laid out by John Goerzen at gopher://gopher.quux.org/0/Archives/Mailing%20Lists/gopher/gopher.2002-02%3f/MBOX-MESSAGE/34
- The UMN format for specifying object attributes and links with .cap, .Links, .abstract, and similar files as specified elsewhere in this document and implemented by UMN gopherd.
- The PYG format for extensible Python gopher objects as created for **PyGopherd**.
- Hypertext Transfer Protocol HTTP/1.0 as specified in RFC1945
- Hypertext Markup Language (HTML) 3.2 and 4.0 Transitional as specified in RFC1866 and RFC2854.
- Maildir as specified in http://www.qmail.org/qmail-manual-html/man5/maildir.html and http://cr.yp.to/proto/maildir.html.
- The mbox mail storage format as specified in http://www.qmail.org/qmail-manual-html/man5/mbox.html
- Registered MIME media types as specified in RFC2048.
- Script execution conforming to both UMN standards as laid out in UMN gopherd(1) and Bucktooth standards as specified at gopher://gopher.floodgap.com:70/0/buck/dbrowse%3ffaquse%202 so far as each can be implemented consistent with secure design principles.

- Standard Python 2.2.1 as implemented on POSIX-compliant systems.

BUGS

Reports of bugs should be sent via e-mail to the **PyGopherd** bug-tracking system (BTS) at pygopherd@bugs.complete.org or submitted on-line using the Web interface at http://bugs.complete.org/. The Web site also lists all current bugs, where you can check their status or contribute to fixing them.

COPYRIGHT

AUTHOR

PyGopherd, its libraries, documentation, and all included files (except where noted) was written by John Goerzen <jgoerzen@complete.org> and copyright is held as stated in the COPYRIGHT section.

Portions of this manual (specifically relating to certian UMN gopherd features and characteristics that PyGopherd emulates) are modified versions of the original gopherd(1) manpage accompanying the UMN gopher distribution. That document is distributed under the same terms as this, and bears the following copyright notices:

PyGopherd may be downloaded, and information found, from its homepage via either Gopher or HTTP:

gopher://quux.org/1/devel/gopher/pygopherd
http://quux.org/devel/gopher/pygopherd

PyGopherd may also be downloaded using Subversion. Additionally, the distributed tar.gz may be updated with a simple "svn update" command; it is ready to go. For information on getting **PyGopherd** with Subversion, please visit:

http://svn.complete.org/

SEE ALSO

python(1).