Pathway Tools Web Server

Why a Web Server?

At least two reasons:

- Publish on your Web site PGDBs that you have created using Pathway Tools, for wide access within your organization, or throughout the Web.
- Access software functionality that is not available in desktop mode, such as some comparative tools available within Pathway Tools. For more details on Web mode versus desktop mode, see `http://biocyc.org/desktop-vs-web-mode.shtml`.

Pathway Tools Web Server

How to Control Ptools Web Server

- Command-line Arguments (User guide section 2.3.1)
- ptools-init.dat (User guide section 2.1)

Command-line arguments override ptools-init.dat.

Pathway Tools Web Server

command line arguments

The following command line arguments are applicable to Pathway Tools only when operating in Web mode:

-www Instructs Pathway Tools to operate in Web server mode.

If ptools-init.dat is configured correctly, this is all you need!

Pathway Tools Web Server

command line arguments

-port NNN This option is valid only when Pathway Tools is operating in Web server mode and is typically used in conjunction with the **-user** option. It specifies in NNN the TCP/IP port on which the Pathway Tools Web server will listen for requests. By default, port 1555 is used on UNIX (Linux), and port 80 is used on Microsoft Windows. See the **-user** argument for UNIX security information.

-user NNN This option is valid only when Pathway Tools is operating in Web server mode and is typically used in conjunction with the **-port** option. It specifies in NNN the UNIX account that Pathway Tools should use to process Web server requests. UNIX allows only **root** accounts to listen on TCP/IP ports numbered up to 1024, so if you specify **-port 80**, for example, then only **root** can start Pathway Tools, but Pathway Tools will switch to running as user NNN right after starting to listen on port **80**. This option is neither available nor necessary on Microsoft Windows.

Pathway Tools Web Server

command line arguments

-proxy-port NNN This argument is valid only when Pathway Tools is operating in Web server mode. It specifies in NNN the TCP/IP port on which another Web server, such as Apache, will listen for requests and forward them to the port on which the Pathway Tools Web server listens for requests. When using this argument, you also have to configure the other Web server, such as Apache, to actually forward the requests. Typical use of this command line argument is to specify **-proxy-port 80** to work around firewall restrictions; most firewalls do allow traffic to port 80.

Pathway Tools Web Server

command line arguments

-www-publish pubspec This argument is applicable when Pathway Tools is running in Web server mode, and affects which of the available PGDBs are visible through the Web server.

Example: Consider a user who wishes to publish all available PGDBs. In this case, use command: **pathway-tools -www -www-publish all**

The full set of possible **pubspec** arguments is

all — make all PGDBs visible

orgA+orgB+...+orgX — make the specified set of organisms visible. Each **orgX** is an organism ID.

public — make public PGDBs visible (excludes the formerly restricted PGDB MetaCyc)

Pathway Tools Web Server

command line arguments

-passwd-file <**file path**> This option allows control over which users can access a given PGDB. The <file path> argument gives the location of the password file in the file system. Once password authentication is turned on, all PGDBs will only be available to authorized users by default. This can be tuned using access control lists, described below for the **-acl-file** option. Note that the usernames defined via this mechanism are completely orthogonal to any user accounts defined via the Web Accounts system described elsewhere.

The format of the password file is simply a series of lines in plain text of the form

user1:password1

user2:password2

etc.

Pathway Tools Web Server

command line arguments

-acl-file <**file path**> Access to PGDBs can be controlled in a fine-grained fashion using Access Control Lists (ACLs). Once passwords are enabled, all published PGDBs are by default restricted to access by authorized users only. If an ACL file is used, each PGDB can have its own access policy. A PGDB can be open (available to all, with no authorization control), or authorized (available only to all users specified in the **-passwd-file** option), or limited to specific authorized users.

This feature is controlled using an ACL file located in the place given by the <file path> argument. The ACL file has the following format:

organism-id :auth

organism-id :open

organism-id user1 user2 user3. . .

Pathway Tools Web Server

command line arguments

Examples:

human :auth

This line would indicate that all authorized users can access the **human** PGDB.

agro :open

This line indicates that anyone, whether authorized or not, can access the **agro** PGDB.

ecoli joe fred tom

This line indicates that only users joe, fred and tom can access the **ecoli** PGDB.

Note that these restrictions prevent any access to the PGDB at all, unless the user is in the authorized group of users. That is, multi-organism queries are also checked to make sure the user is only allowed to access that set of organisms for which he is authorized.

Pathway Tools Web Server

command line arguments

To facilitate the entry of long lists of users, the ~ character can serve as a line-continuation character.

Example:

```
ecoli      ~
     joe   ~
     fred  ~
     tom
```

would have the same effect as the line

```
ecoli    joe fred tom
```

Pathway Tools Web Server

command line arguments

-gene-link-db db When a user site sets up a Pathway Tools database in conjunction with a previously existing database of genes for an organism, it is sometimes useful to have references to genes in Pathway Tools Web pages link directly to gene pages in the user's pre-existing database, rather than to the gene pages generated by the Pathway/Genome Navigator. To accomplish this, the user must create a Database frame in the PGDB that contains information necessary for linking to the desired external database, and each gene frame in the PGDB must contain a link to the corresponding object in that database. Then, if the Database frame ID is supplied as the value of this command line argument, pages generated by the Pathway Tools Web server will substitute links to the external database anywhere it would normally link to a gene page.

Pathway Tools Web Server

command line arguments

-no-blast Ordinarily the Web server's Query Page contains an option to invoke the BLAST program. If access to BLAST from this page is not desired (e.g., if the BLAST program is not installed, or if such functionality is available elsewhere on a Web site), then supplying the **-no-blast** option causes it to be removed from the query page. Note that BLAST is disabled by default when the **-gene-link-db** argument is supplied.

-blast Restore access to the BLAST program from the Web query page if it was removed by virtue of the **-no-blast** or **-gene-link-db** arguments being supplied.

Pathway Tools Web Server

command line arguments

-no-google-text-search By default, if your Web server is served from port 80, the Pathway Tools query page will contain a full-text-search query box, powered by Google™. This capability is useful only if the site can be indexed by Google. If your Web server is running on an internal network or cannot be indexed by Google for other reasons, supply this argument to remove the search box.

-google-text-search By default, if your Web server is not served from port 80, the Pathway Tools query page will not contain a full-text-search query box, powered by Google™. If your Web server is indexed by Google nevertheless, and you want the search box to appear, supply this argument.

-email support@site Specifies the email address to which technical and content-related support questions should be addressed. This address will appear on the Pathway Tools Web pages.

Pathway Tools Web Server

command line arguments

-allow-webcrawlers The Web server mode activates Webcrawler detection by default. When too many requests come in from the net, which originate from the same IP number within a certain time period, such an IP number will be blocked from further access, until the Pathway Tools server is rebooted. The assumption is that aggressive Webcrawlers (from spammers) are taxing the performance from the Web server substantially, making access for *bona fide* users more sluggish. However, under certain circumstances, it may be desirable to bypass the Webcrawler detection, which can be achieved by supplying this argument, when calling the **pathway-tools** script.

Pathway Tools Web Server

ptools-init.dat

```
# Generated by SRI International Pathway Tools version 12.0 on Wed Aug 27, 2008
##########################################################################
#
#    Pathway Tools Init File
#    -----------------------
#    This file contains initialization parameters for Pathway Tools.
#    Each parameter needs to be specified on its own line, with the name
#    of the parameter followed by a separator consisting of one or more space characters,
#    which are followed by the value, which is usually a text string or an integer.
#
#    Lines beginning with the hash character "#" are ignored and are used for comments.
#    Therefore, when defining a value for a parameter whose line begins with "#",
#    please be sure to remove all hash characters from the line for that parameter
#    or it will remain inactive.
#
#    The parameters in this file can usually be overridden by supplying the
#    corresponding command-line flags to the pathway-tools startup script.
#
##########################################################################
```

Pathway Tools Web Server

ptools-init.dat

```
##########################################################################
#
#    Parameters for Pathway Tools when running in WWW server mode
#      -- Setting these parameters is MANDATORY if you run Pathway Tools in
#         WWW server mode.

# Hostname of the machine on which Pathway Tools is running,
# which Pathway Tools will insert in URLs within the WWW pages
# that it generates.
#
WWW-Server-Hostname localhost

# Port number on which Pathway Tools should listen for WWW server
# requests, and which Pathway Tools will insert in URLs within the
# WWW pages that it generates.  The standard port for WWW servers is 80,
# but this may require root privileges for starting up Pathway Tools.
#
###WWW-Server-Port 1555
```

Pathway Tools Web Server

ptools-init.dat

```
##########################################################################
#
#    Additional Parameters for Pathway Tools when running in WWW server mode
#      -- Setting these parameters is OPTIONAL if you run Pathway Tools in
#         WWW server mode.

# The email address to which users of your Web site should send technical-support
# email messages.  This address will be inserted into pages generated on your Web site.
#
###Support-Email-Address biocyc-support@ai.sri.com

# The name used for the Web site being operated by Pathway Tools.
# This name will be inserted by Pathway Tools into pages that it generates
# on your Web site.
#
###WWW-Site-Name BioCyc
```

Pathway Tools Web Server

ptools-init.dat

```
# Specify which organisms should be published by the WWW server.
# This parameter assists users in complying with the SRI license
# provision that disallows users from publishing the MetaCyc PGDB
# on their externally visible WWW site.  However, users may publish
# all PGDBs on a site only visible from within their institution.
# Possible values are public, all, or a list of organism IDs delimited by +'s.
# Example of the latter: ECOLI+META+VCHO
#
###WWW-Publish public

# Needed only for a Proxy WWW Server setup.
# Port number on which another WWW server, such as Apache, will listen
# for requests to forward them to the Pathway Tools WWW server.
# When using this parameter, the other WWW server needs to be configured
# to actually forward the requests.
#
###WWW-Server-Proxy-Port 80
```

Pathway Tools Web Server

ptools-init.dat

```
# If a WWW-Server-Port smaller than 1024 is specified, on UNIX, the root user
# needs to start the Pathway Tools, and after the WWW server starts listening,
# it will switch to the user account indicated here, to reduce security risks.
#
###WWW-Server-User biocyc
```

Pathway Tools Web Server

ptools-init.dat

```
# If WWW-Quick-Search-Textfield-Label is set to a given string, then that string
# will appear to the left of the Quick Search text box at the bottom of every
# database page. The default is to have it say, "Quick Search".
# To enter spaces into the string, use the underscore key: "Quick_Search".
#
###WWW-Quick-Search-Textfield-Label Quick_Search

# Set to "Y" to have the Query Page display the "Summary page for dataset"
# link. A value of "N" suppresses it.
#
###WWW-Show-Organism-Summary-Link Y

# Set to "Y" to have the Query Page display the "History of updates for this
# dataset" link. A value of "N" suppresses it.
#
###WWW-Show-Update-History-Link Y
```

Pathway Tools Web Server

ptools-init.dat

```
# Set to "Y" to have the Query Page display the "Cellular Overview Diagram /
# Omics Viewer" link. A value of "N" suppresses it.
#
###WWW-Show-Diagram/Omics-Viewer-Links Y

# Set to "Y" to have pathway displays include predicted enzymes by default.
# A value of "N" means that only enzymes with experimental evidence will
# appear unless the user requests to see all enzymes.
#
###WWW-Default-Show-Predicted-Enzymes Y
```

Pathway Tools Web Server

ptools-init.dat

```
##########################################################################
#
#    Parameters used by Pathway Tools when accessing PGDBs within a MySQL
#    or Oracle server.
#      -- Setting these parameters is MANDATORY if you want to access PGDBs
#         from a MySQL or Oracle server.

# Hostname of the machine on which the MySQL or Oracle PGDB server is running.
#
###RDBMS-Server-Hostname

# Port number on which either the MySQL or Oracle PGDB server is listening.
# MySQL standard is 3306; Oracle standard is 1521
#
###RDBMS-Server-Port 3306
```

Pathway Tools Web Server

ptools-init.dat

```
# Name of database within the MySQL or Oracle server in which PGDBs
# should be stored.
#
###RDBMS-Database-Name

# Username used to log in to the MySQL or Oracle server.
# If set to the value PROMPT, then the user will be prompted
# for their username and password.
#
###RDBMS-Username

# Password used to log into the MySQL or Oracle server.
#
###RDBMS-Password
```

Pathway Tools Web Server

ptools-init.dat

Note: 12.5 feature

```
##########################################################################
#
#    Parameters for the Web Accounts system of  Pathway Tools
#      when running in WWW server mode
#      -- Setting these parameters is OPTIONAL if you run Pathway Tools in
#         WWW server mode.
#
# The Web Accounts system is used to manage logins, user profiles, and user
# preferences.
#
# Hostname of the machine on which the MySQL Web Accounts server is running.
#
###User-Account-RDBMS-Server-Hostname

# Port number on which the MySQL Web Accounts server is listening.
# MySQL standard is 3306
#
###User-Account-RDBMS-Server-Port 3306
```

Pathway Tools Web Server

ptools-init.dat

```
# Name of database within the MySQL Web Accounts server in which user
# information should be stored.
#
###User-Account-RDBMS-Database-Name

# Username used to log in to the MySQL Web Accounts server.
#
###User-Account-RDBMS-Username

# Password used to log into the MySQL Web Accounts server.
#
###User-Account-RDBMS-Password
```

Pathway Tools Web Server

ptools-init.dat

Note: 12.5 feature

```
# The ReCaptcha system is used  by Web Accounts to authorize submissions
# by humans but block web-bots. This is the ReCaptcha private key string
# tied to your server domain yourserver.com.

#
###User-Account-ReCaptcha-private-key

# This is the ReCaptcha public key string, also tied to your server
# domain yourserver.com. You can get your own private and public keys for
# your domain for free from www.recaptcha.net.
#
###User-Account-ReCaptcha-public-key
```

Pathway Tools Web Server

Misc. Considerations

Running Unattended At SRI we use **screen** + **Xvfb** + **emacs**. See `http://bioinformatics.ai.sri.com/ptools/web-logout.html`.

Bulk Loading of DBLinks DBLinks allow you to link PGDB objects to foreign databases. Two files:

- Database definition file (for foreign databases). Specified with **-dbdef** command-line argument.
- Link definition file (to show relationships between objects in PGDB and foreign database. Specified with **-linkdef** command line argument.

See User Manual, section 7.5.8.

Much More Available in User Manual Customization, troubleshooting, etc. etc.