cPath Administrator Guide

Introduction

This document describes the procedure for installing and administering cPath.

System Requirements. To install cPath, you must have the following installed locally:

- MySQL: Information available at: http://www.mysql.com/.
- Apache Tomcat Server, 4.1X or above: Information available at: http://jakarta.apache.org/tomcat/ Note that cPath has not (yet) been tested on Apache Tomcat, 5.X or above.
- Apache Ant: Information available at: http://ant.apache.org/.
- Perl 5.0 (or later): Required for running the cPath Administrator Command Line Tool.

cPath Installation: Step-By-Step Instructions

1. If you have not already done so, install MySQL, Apache Tomcat, and Ant locally. Once you have verified that the installations are working correctly, proceed to step 2.
2. Create a mysql user for cPath. For easiest installation, use the following:

   user: tomcat@localhost
   password: kitty

If you choose a different user/password combination, see Appendix A below.

Instructions on adding new mySQL users is available online at: http://www.mysql.com/doc/en/Adding_users.html.

Make sure this user has permission to update the database.

3. Load the cPath database tables into MySQL. To do so:

```
cd [CPATH_ROOT]/dbData
mysql -u tomcat -p < cpath.sql
```

You will be prompted for a password. Enter the value: "kitty".

4. Load a Demo Set of Sample data

```
cd [CPATH_ROOT]/dbData
mysql -u tomcat -p < demo.sql
```

You will again be prompted for a password. Enter the value: "kitty".

5. Copy dist/cpath.war to [JAKARTA_TOMCAT_DIR]/webapps and restart Apache Tomcat.

   Note: On certain platforms, you may not even need to restart Tomcat. For example, if you are running Tomcat 5 on Windows, Tomcat will immediately detect the new WAR file, and load it automatically.

6. Open a web browser, and go to: http://localhost:8080/cpath

   Verify that you can see the cPath Home Page.

7. As a final step, you must run the full text indexer. The full text indexer indexes all entities in cPath, and makes those entities available via the cPath search box.

   - Go to: http://localhost:8080/cpath/adminHome.do
   - When prompted, enter the following:

   user: admin
   password: cpath

   - Select the link: "Run Full Text Indexer" -- full indexing of the demo set takes approximately 30 seconds.

cPath Administration

You can administer your instance of cPath via a web interface or via a command line interface. The web-based administrator is easy to use, but you may find the command line interface is easier for long running tasks. For example, if you are importing a very large PSI-MI file, or re-running the full-text indexer on a large set of data, it is easier to do so via the command line tool. The command line tool also supports a few more options, and currently provides more detailed error and warning messages.

Administering cPath via the Web

To access the cPath Administration Web Page:

- Go to: http://localhost:8080/cpath/adminHome.do
- When prompted, enter the following:
  user: admin
  password: cpath

From this page, you can do the following:

- Run cPath Diagnostics: This page verifies that the MySQL connection is working properly, and that all the correct cPath tables are in place. It also verifies that the full-text search engine is set up correctly.
- Run the full-text indexer: The full text indexer indexes all entities in cPath, and makes those entities available via the cPath search box.
- Import PSI-MI Data Files: Imports new PSI-MI data files into cPath.
- Toggle Web Diagnostics: Each time cPath generates a web page, it can generate real-time web diagnostics, which are appended to the bottom of the page. You can activate / deactivate web diagnostics here.
- Reset Organism Stats: To optimize performance, organisms stats are cached locally in memory, and are used to build the “Browse By Organism” page. If you run the full-text indexer from the web interface, the organism stats are automatically reset. However, if you run the full-text indexer from the command line, you may need to reset the organism stats manually from this page.

Note: When you run the full-text indexer via the web interface, existing index files are automatically deleted and the new index files are created in-place at: /cpath/build/WEB-INF/textIndex/. If a user executes a search during a long-running index operation, the search will work correctly, but the user will not likely see a complete set of matching search results. If you want to avoid this type of problem, run the full text indexer from the command line (see details below.) The command line tool does not modify existing index files used by end-users. Rather, it creates index files in a separate local directory. When indexing is complete, you can copy the complete local index to your web application. This ensures that end-users always receive a complete set of accurate results.

Administering cPath via the Command Line

The cPath Command line tool is available in bin/admin.pl. Here is a summary of the command line usage:

```
cPath Admin.  cPath Version:  0.3.1 (Beta)
Copyright (c) 2004 Memorial Sloan-Kettering Cancer Center.

Administration Program for the cPath Database
Usage:  admin.pl [OPTIONS] command
  -f, -f=filename Name of File / Directory
  -d,             Shows all Debug/Log Messages/Stack Traces
  -u, -u=name     Database User Name
  -p, -p=name     Database Password
  -x              Skips Validation of External References
  -o, -o=id       NCBI TaxonomyID

Where command is a one of:
  import          Imports Specified File.
                  Used to Import PSI-MI Files or External Reference Files.
  index           Indexes All Items in cPath
  precompute      Precomputes all queries in specified config file.
  count_affy      Counts Records with Affymetrix identifiers.

Extra Options (not guaranteed to be available in future versions of cPath)
  -r              Removes all Interaction PSI-MI xrefs  (not recommended)
                  Used to temporarily import buggy HRPD PSI-MI Files.
```

The two most common commands are “import” and “index”. Details on these two commands are provided below.

Importing New PSI-MI Data Files

To import new data into cPath, the data must be specified in valid PSI-MI XML format. Details regarding PSI-MI are available online at: http://psidev.sourceforge.net/. To get you started, a number of sample PSI-MI files are available under the /testData directory.

To import new PSI-MI data, use admin.pl with the import command. For example, the following command will import one of the sample data files in /testData:

```
./admin.pl -u tomcat -p kitty -f ../testData/dip_psi.xml import
```

The admin tool will report progress, errors and warnings to the console.

Running the Full Text Indexer

Once you have imported new data into cPath, you need to re-run the full-text indexer. As stated above, the full text indexer indexes all entities in cPath, and makes those entities available via the cPath search box.

Here is an example of how to run the full-text indexer from the command line:

```
./admin.pl -u tomcat -p kitty index
```

This will automatically generate new index files in: /cpath/build/WEB-INF/textIndex/. After indexing is complete, you will need to copy these files to your webapp directory, located at: /webapps/cpath/WEB-INF/textIndex.

A sample Perl script that we use locally to automatically deploy new index files is available at bin/deployText.pl. This script is hard-coded for our production environment, but you can use it as a template.

Appendix A: Creating a cPath MySQL User

If you choose a MySQL user other than user=tomcat/password=kitty, you must update two configuration files:

- [CPATH_ROOT]/config/config-JDBC.properties; and
- [CPATH_ROOT]/web/WEB-INF/web.xml

Replace the words tomcat/kitty with your own username/password.

Appendix B: Monitoring cPath for Constant Uptime

Once installed, it is useful to monitor cPath on a periodic basis, and verify that it is running correctly. A recommended tool is TinyMonitor, available from: http://www.glug.com/projects/TinyMonitor/. From the Glug.com site:

> "TinyMonitor was written out of pure necessity for a simple monitoring program that watched the actual content of returned pages rather than simply checking to see if the httpd service was running. (Experience has

proven that the web server running doesn't necessarily mean it's reporting the content you want.)”

This program can be used for simple web server monitoring (i.e. is it actually delivering content?) or can be used to verify a page is returning what you expect (i.e. a 200 rather than a 404). This is an excellent choice for someone who doesn't want to spend thousands on a program that will do actual http content monitoring (i.e. SiteScope or HP OpenView Internet Services)."

A copy of TinyMonitor (along with a sample config file for monitoring cPath) is available in bin/tinymonitor.

Appendix C: Where to Look When Things Go Wrong

If you experience difficulty with cPath or notice unusual errors, there are two places you should look first:

- Real-time Web Diagnostics: Each time cPath generates a web page, it can generate real-time web diagnostics, which are appended to the bottom of the page. To activate web diagnostics, go to the Admin page, and click the "Activate" link under Web Diagnostics. If the page generates an error, the web diagnostics section will include detailed error information, which would not normally be shown. This may help track down the source of the error.

- Servlet Engine Logs: In the event of an error, cPath will also write the error to the standard servlet engine log. For example, when using Apache Tomcat, error messages are written to catalina.out.

Appendix D: Troubleshooting

Known Issue #1: PacketTooBigException

Symptoms: You attempt to import a very large data file, and receive the following error message: PacketTooBigException.

Diagnosis: This error occurs because cPath stores the complete contents of all imported records in the MySQL database. If the imported record exceeds the MySQL max_allowed_packet setting, you will receive a PacketTooBigException.

Solution: To fix this problem, increase the MySQL setting for max_allowed_packet.

The simplest way to do so is to update (or create) a my.cnf configuration file. For example, the following file (/etc/my.snf) increases the max_allowed_packet to 65M:

```
[mysqld]
max_allowed_packet=65M
```

Resources:

- Complete details regarding the "Packet too Large" issue is available at: http://dev.mysql.com/doc/mysql/en/Packet_too_large.html.
- Complete details regarding "Using Option Files", and configuring MySQL settings, see: http://dev.mysql.com/doc/mysql/en/Option_files.html.

Known Issue #2: OutofMemoryError

Symptoms: You attempt to import a very large data file via the cPath Administrator web page, and receive the following error message: OutOfMemoryError

Diagnosis: This error occurs because cPath stores the complete contents of a single imported record in memory before writing to the MySQL database.

Solution: To fix this problem, increase the memory available to Tomcat and restart. The simplest way to do so is to update the $TOMCAT_HOME/bin/cataline.sh file (for Tomcat versions 4.x)
You can specify '-Xmx512M' as part of JAVA_OPTS. That will set your max JVM heap size to 512MBs, for example.