cPath Administrator Guide

Revisions

Date	Description
November 16, 2005	Updated to reflect changes in cPath, version 0.5. Now includes details re: revised installation procedure, BioPAX v. PSI-MI, use of build.properties, etc.

Introduction

This document describes the procedure for installing and administering cPath.

Note: This Administrator Guide applies to cPath, Version 0.5 and later only.

System Requirements

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

- MySQL: Information available at: http://www.mysql.com/. cPath has been tested with MySQL 4.0 and 4.1.

- Apache Tomcat Server, 4.1X or above: Information available at: http://jakarta.apache.org/tomcat/. cPath has been tested with Tomcat 4.1 and Tomcat 5.0.
- Apache Ant 1.6 (or later): Information available at: http://ant.apache.org/.
- Perl 5.0 (or later): Information available at: http://www.perl.org/.

cPath Limitations: PSI-MI v. BioPAX

As of Version 0.5, cPath is capable of storing records in one of two formats: PSI-MI or BioPAX. This means that you can load cPath with all PSI-MI data or all BioPAX data, but cannot load a mix of the two. The default installation, and the instructions below assume a BioPAX installation only.

cPath Installation: Step-By-Step Instructions

1. Install required software: 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. Install Third-Party Libraries Required by Ant: The ant build.xml file requires a few third-party libraries. Copy all .jar files in [CPATH_HOME]/lib/ant to [ANT_HOME]/lib. For example:

```
cp lib/ant/*.jar /Users/cerami/libraries/apache-ant-1.6.1/lib
```

3. (Optional) Update your build.properties file: Here is an excerpt from the build.properties file:

```
# Application Name and Version
app.name=cpath
app.version=0.5

# Database Settings
db.name=cpath
db.user=tomcat
db.password=kitty
db.host=localhost
```

By default, the database name is “cpath”, and we have provided a sample user to get started. For the initial installation, it’s easiest to just keep this file as is.

4. Create a mysql user for cPath: For easiest installation, use the following username / password combination.

user: tomcat@localhost
password: kitty

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.

5. Set a CPATH_HOME Variable: In order to proceed, you will need to set a CPATH_HOME environment variable. If you are using Mac OS X or Linux, you can do so via the shell. For example, here is an excerpt from a .bash_login file:

```
# Set up CPATH_HOME Variable
export CPATH_HOME=/Users/cerami/dev/sander/cpath
```

6. Create the cPath Database in MySQL: To do so, run the initDb.pl script in the bin directory. For example:

```
> ./initDb.pl
Script to Initialize cPath Database
===================================

Using Database Name:      cpath
Using Database Host:      localhost
Using Database User:      tomcat
Using Database Password:  kitty

! Running this command will delete all existing data !
Are you sure you wish to proceed (Type: YES):  YES
Importing MySQL File:  cpath.sql
Importing MySQL File:  seed.sql
Done.  cPath is now ready to go...
```

7. Compile the Code: To do so:

```
> cd [CPATH_HOME]
> ant
```

8. Load a Sample BioPAX File: To do so, run the admin.pl script in the bin directory. For example, the following command loads the Apoptosis pathway from Reactome:

```
> ./admin.pl -f ../testData/biopax/Apoptosis.owl import
cPath Admin.  cPath Version:  0.5 (Beta)
Copyright (c) 2005 Memorial Sloan-Kettering Cancer Center.

---------------------------------
Database Name:     cpath
Database Host:     localhost
Database User:     tomcat
```

```
Database Password: kitty
----------------------------------

XML Type:  BIO_PAX
Importing File...
Validating BioPAX File with RDF Validator...
   --->  BioPAX RDF is Valid
Categorizing BioPAX Resources
Validating RDF Links
…
```

9. Generate a Full-Text Index: This is required for the cPath search feature. To do so, run the admin.pl script in the bin directory. For example:

```
>./admin.pl index
cPath Admin.  cPath Version:  0.5 (Beta)
Copyright (c) 2005 Memorial Sloan-Kettering Cancer Center.

----------------------------------
Database Name:     cpath
Database Host:     localhost
Database User:     tomcat
Database Password: kitty
----------------------------------

Running Full Text Indexer
Clearing XML Cache
Indexing all cPath Records
Total Number of Pathways:  32
Total Number of Interactions:  81
Total Number of Physical Entities:  81
…
```

10. Generate the cPath WAR File: The WAR (Web Application Archive) contains the complete contents of cPath, ready for deployment to Tomcat. To generate the war file, go to CPATH_HOME, and run:

```
ant war
```

11. Deploy cpath.war to Tomcat: To do so, simply copy the cpath WAR file from build/war to your Tomcat webapps directory. For example:

```
cp build/war/cpath_0.5.war ~/libraries/jakarta-tomcat-5.0.28/webapps/
```

Copying your war file here will cause Tomcat to automatically load and start cPath.

12. Verify Installation: Open a web browser, and go to: http://localhost:8080/cpath_0.5/. Verify that you can see the cPath Home Page.

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 files, 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 the cPath Home Page: http://localhost:8080/cpath_XX/.
- Click on the Admin Tab.

  Note that this tab will only appear on a local installation of cPath. If you are running on a non-local installation, e.g. in production mode, you need to type the URL is directly: http://localhost:8080/cpath_XX/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 Data Files: Imports new PSI-MI / BioPAX 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.
- Configure the Web UI: For example, you can configure a header, footer, and home page content.

Note: If you wish to update the admin username / password, update build.properties, and redeploy your WAR file.

Note: If you wish to disable the web admin functionality altogether, update build.properties to: web.admin.active=0, and redeploy your WAR file.

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:  0.5 (Beta)
Copyright (c) 2005 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 (overrides build.properties)
  -p, -p=name     Database Password (overrides build.properties)
  -h, -h=hostname Database Server Name (overrides build.properties)
  -b, -b=database Database name (overrides build.properties)
  -x              Skips Validation of External References
  -o, -o=id       NCBI TaxonomyID
  -q, -q=term     Full Text Query Term

Where command is a one of:
  import          Imports Specified File.
                  Used to Import BioPAX Files, PSI-MI Files
                  ID Mapping Files, or External Database files
  index           Indexes All Items in cPath
  precompute      Precomputes all queries in specified config file.
  count_affy      Counts Records with Affymetrix identifiers.
  validate        Validates the specified XML file.
  query           Executes Full Text Query
```

Working with PSI-MI Data

As noted above, you can load cPath with all PSI-MI data or all BioPAX data, but cannot load a mix of the two. If you wish to load PSI-MI data, there are three steps:

- Clear out the existing database of all data. To do so, just run: bin/initDb.pl.
- Load your PSI-MI data via bin/admin.pl
- Update build.properties to: web.mode=psi-mi. This causes the web front-end to look like earlier versions of cPath, and hides all BioPAX related functionality.

Appendix A: 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 B: 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 C: 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.