Chemaztech handbook

Summary

Chemaztech handbook....1
Introduction....2
Features....2
Installation....4
Basic Install....4
LAMP Server....4
OpenBabel....4
MyChem....5
JRE....5
Installation of OASA library, libcairo and python....5
Plugins Installations....7
TCPDF....7
PHPExcel....7
ChemAzTech Installation....7
Automatic installation (since version 0.6)....9
Before version 0.6....10
Settings....11
Config.php....11
config_ldap.php....11
First Access....12
PHPIDS....12
Web Server and PHP....12
php.ini....12
Suhosin....13
Using....14
General patterns and overall functionning....14
Frequently Asked Questions....17
Directory tree....17
- MyChem Problems....17
Notes and references....18
Appendix....19

Introduction

Currently, with the exception of a few products, tools for chemistry remain largely for a single "chemist" (or they are extremely expensive).

The purpose of this software is to provide a chemical library, a collaborative tool for managing all the products of a laboratory easily from a browser.
The software is completely free, under "CeCILL french license", or the famous "GNU GPL" if you are elsewhere.

The chemical tool is based on PHP and MySQL, but it includes/needs other specific chemical tools :

- OpenBabel,
- MyChem,
- JchemPaint,
- Jmol,
- libcairo

A little bit of python and some extraction tools (XLS(X)/PDF) complete the software.

I must admit this is a long list, but do not forget it is a free software !

Features

- Management of chemistry product (with structures, storage, references, experiment number...),
- Draw, edit structures with Jchempaint,
- View molecules in 3D with Jmol, in 2D with JCP (java) or OASA python library (SVG picture),
- Substructures or similarity search with MyChem and a lot of additionnal features from MyChem,
- Import molecules from MOL/SDF/CDX files,
- CSV/XLS(X) extraction, PDF exports...,
- Manage Users and Groups.

Important Note : Viewing pictures from « libcairo » library is sometimes wrong. The *java* display, is correct every time, but longer to generate. I look forward to an update of this library. However, SVG display is more precise than PNG (in the previous version).

Advanced using for High Performance Computing

Since some recent release, I recall the possibility of using a parallel cluster to test molecules *in silico.* On this point, I hope you have a little bit of developer soul, because it is not entirely easy.

For my part, I gave the possibility to send a molecule to a computing cluster in a target directory with structures (that you set previously[1]). Sending and extra commands are set in the general configuration, in “config/config.php” file. Then the page “visu3D.php” in “jmol/mol” directory call these commands.

It is also possible to do the opposite : manually export the whole database for virtual screening *in silico*, but it is not easy. Script *sdf_correct.py* can help you in this process by putting the names inside the molecules with the experiment numbers (known in the database). First, you have to export a SDF file from the database, with the admin_panel interface.

1 For my part I use a cron script on the master node to see if new molecules have been received recently.

The installation is what requires the most effort, like a lot of software... There are several facilities to achieve a functional system. It is a really disadvantage.

First, you must install a LAMP server up to date, that is to say, a linux server that contains the latest versions of MySQL, PHP and Apache. The, you have to install and configure the JRE from java sun/oracle, then “OpenBabel”, “MyChem”, “python” and finally the “libcairo” library.

With all these things, you will be almost ready :
Exporting in XLS and PDF require 2 plugins :
- PHPExcel : http://phpexcel.codeplex.com/
- TCPDF : http://www.tcpdf.org/

Basic Install

LAMP Server

I believe that if you want to install my software, you know how to install a LAMP web server, or that you already have one.

With a linux OS - debian ~like :

```
$sudo apt-get install apache2 apache2-doc mysql-server php5 libapache2-mod-php5 php5-mysql phpmyadmin
```

Also note that Ubuntu also offers a server version with automatic LAMP installation. You could do the same thing with another package manager : *apt*, *yum* or *urpmi*...

```
#yum install httpd mysql mysql-server php phpmyadmin
```

NB : PHPMyAdmin is not absolutely required but it is very usefull...
NB2 : Vérify the name of the packages; the name is changing with new release.

If you do so (above), installation ask to configure the Apache server and the MySQL database. Do not forget the root password of your database !

OpenBabel

OpenBabel is also generally present in the package managers like apt or synaptics. However, check that it is at least version 2.2.0.

```
#apt-get install openbabel
```

Otherwise, you can install it from the source (*configure, make, make install* ...) :
http://openbabel.org/wiki/Install

MyChem

MyChem, is more difficult to install...

I suggest you to read the associated documentation:
http://mychem.sourceforge.net/doc/apas03.html

In case of trouble with compiling, please check your version of gcc and g++, and switch to verbose mode.

```
#make VERBOSE=1
```

Note

Chemaztech uses many functions from MyChem. If it is not installed correctly, the application will only serve to make management and cannot calculate any chemical property and do not allow searching by substructure or similarity...

Without OpenBabel, MyChem cannot work. Moreover OpenBabel is used directly by -about- two scripts (including file conversion CDX MOL), even if I limited the usage of external system call (babel ...).

JRE

You have to follow instructions from Sun Website :
http://java.com/fr/download/help/linux_install.xml

JDK is not required.

Installation of OASA library, libcairo and python

```
#aptitude install python-central avec libcairo et python-cairo
```

Source of OASA library can be found on BKChem Website :
http://bkchem.zirael.org/oasa_en.html

OASA library needs, at least, python version 2.3 :

```
#python setup.py install
#python
```

```
>>>import oasa
>>>exit()
```

Now, your system can create SVG or PNG pictures, or PDF files from a mol data with python.

Plugins Installations

TCPDF

Without this plugin, PDF generation will not work. To install it, you have to follow this : http://www.tecnick.com/public/code/cp_dpage.php?aiocp_dp=tcpdf_installation

To resume, you have to put these files into a directory at your Web Document Root on your server. Then, you have to edit the configuration file : `config/tcpdf_config.php`

Be sure to give *Apache* user enough access to the directory (644) :

```
#chmod -R 644 <directory_name>
```

PHPExcel

PHPExcel plugin allow XLS and XLSX creation from PHP. If you do not install it on your server, users could only export CSV files from their database extractions (in CSV file, you can not have product's 2D structure). PHPExcel's license is LGPL. The installation is easy (like Tcpdf). However, PHP 5.2 is required and your server must have some PHP plugins activated :

- php_zip
- php_xml
- php_gd2

Cf : http://phpexcel.codeplex.com/wikipage?title=Requirements&referringTitle=Home

Then, you have to create a directory at the Web Document Root and copy into it the content of *« Classes »* directory (or you can copy *Classes* directory to the Web Document Root).

Once again, be sure to give Apache user enough access to the directory (read + write).

ChemAzTech Installation

Installation of ChemAzTech is very simple : unzip the archive to the Web Document Root and give enough rights to *Apache* user.

```
#cd $DOCUMENT_ROOT (e.g. : cd /var/www )
#tar -xzvf chimiotheque.tar.gz
#chown -R apache_user chimiotheque
```

Debian => *apache_user* = *« www-data »*

NB : *import*, *export, zip, pdf/** and *jmol/mol* directories must be writeable.

```
#chmod -R ug+rx chimiotheque
#chmod -R u+w chimiotheque
```

Automatic installation (since version 0.6)

This paragraph explains how to configure and install Chemaztech software. If your version of Chemaztech is older than the version 0.6, thank you to ignore this text.

Go to the following address :

http://[ChemAzTech Web document root]/config/install.php

Follow all the steps and you will have a basic functional installation. You can also configure a LDAP authentication and choose the type of users they will be : temporary user or simple user (they do not have the same rights).

(See "Using"> "General diagrams and overall operation"> I] rights)

login >

user : admin

password : admin

Once installation is complete, log into admin to configure software (see "Configuration"> "First Access" for details.)

However, some options need to be edited directly from the config file on the server for security reasons. I advise you to read « Settings » in this documentation to change your general options and to modify *Suhosin, PHPIDS* and *PHP.INI*.

Before version 0.6

Manual configuration of database :

Execute the good SQL file in « config » directory depending if you do a new install or an update of the program. In config directory *chimiotheque*Struct.sql* contain database's structure; it is for a new install. For an update, execute *update*.sql* corresponding to your version.

```
#mysql < *your_file*.sql
```

Config.php

Most parameters could be initialized in *config.php* file from *config* directory.

First lines contains informations to connect to the database :

```
$host = "";
$db = "";
$user = "";
$pass = "";
```

Usually $db = « chimiotheque », unless you have modified the database's name. If MySQL is installed on the same server $host= « localhost ». At last, you have to create a user and a password to use the database (for example, you can do it with PHPMyAdmin) (give him full access to the *"chimiotheque"* database). This is a typical LAMP software installation. If you do not use PHPMyAdmin, you can also do this :

```
# mysql -u root -p
mysql> grant all privileges on chimiotheque.* to
your_chimiotheque_user@localhost identified by
'your_password_for_chimiotheque_db';
```

The next two steps in the configuration of this file is for **TCPDF** and **PHPExcel**. You should activate them to enable these plugins :

$phpexcel and $tcpdf should be set to « 1 ».

To disable these two plugins, set these values to « 0 ».

If you enable the plugin you have to put the path in the config file. Define it from ChemAzTech root directory; For example, if the directory of PHPExecl is call *Classes* at your Web Document Root ($DOCUMENT_ROOT), the relative path to PHPExcel will be « ../Classes/ ».

```
$tcpdf = 1;
$tcpdf_path = "../tcpdf/";
$phpexcel = 1;
$phpexcel_path = "../Classes/"
```

The next step is to insert the mail of the product manager. He will be warn if a product is empty (quantity = 0).

Finally, the last thing you can do in this file is to configure a cluster and a cluster manager (define an e-mail) (*it is complicated; See* ***Advanced using for High Performance Computing*** *in introduction*).

config_ldap.php

This file contain informations about LDAP connection, to authenticate your users if you are using a directory. Since version 0.6, you can modify directly these informations from the *administrative (tasks)* panel « configure ChemAzTech options ». You can choose what kind of access these users will have for their first access : simple user or temporary user (they do not have the same rights).

(See "Using"> "General diagrams and overall operation"> I] rights)

First Access

To login, you just have to go to "Login(out)" in the top menu, and enter your username and password :

User : admin *password* : admin

Do not forget to change the default password at the first login. Go to *"Administrative tasks" > "Modify my personal informations" > "Change password"*.

PHPIDS

If you see alert messages like "Danger detected Bye Bye", you have to change the PHPIDS detection level. You can read *«Frequently Asked Question » > « About PHPIDS ».*

PHPIDS, and *HTMLpurifier* allow to prevent intrusions from PHP attacks. A low score (= equivalent to a high level of security) do not permit to use the software properly, especially if you have a big amount of molecules.[2]

Web Server and PHP

Even if I consider that you are able to do that, I have to specify some points, particularly about the timeout or file size limit.

php.ini

php.ini values are too low. If you want to upload a 3Mo CDX file to server, or import a 7Mo SDF file, you have to change this :

Under debian distro : `# vim /etc/php5/apache2/php.ini`

2 PHPIDS detected score depending of the total amount of variables. Even if the danger level is "5" by variable (it is a low security level), but if you have 300 variables to confirm to next page (e.g. *via* the sript *choix.php* which permit to choose the molecules following a search request), the page score will be the total (300*5)...

```
output_buffering = 4096
max_execution_time = 1000        ; Maximum execution time of each
script, in seconds
max_input_time = 600 ; Maximum amount of time each script may spend
parsing request data
memory_limit = 1024M        ; Maximum amount of memory a script may
consume (128MB)
register_globals = On
post_max_size = 256M
enable_dl = On
upload_max_filesize = 16M
default_socket_timeout = 600
auto_detect_line_endings = On
extension=mysql.so
extension=gd.so
extension=fileinfo.so
mysql.connect_timeout = 600
gd.jpeg_ignore_warning = 0
```

Above, I have listed all the default values that I modified. Those which interest us are : **upload_max_filesize**, **max_execution_time**, **post_max_size** et **memory_limit**. I have choosen high values, because of the high quantity of products in our database and the significant size of file to import (SDF), but choose the values that corresponding to your products.

If you want to get more informations about this, read : http://php.net/manual/fr/ini.core.php

Suhosin

Note that you could also have problems from ***suhosin*** with these limit values :

From Debian :

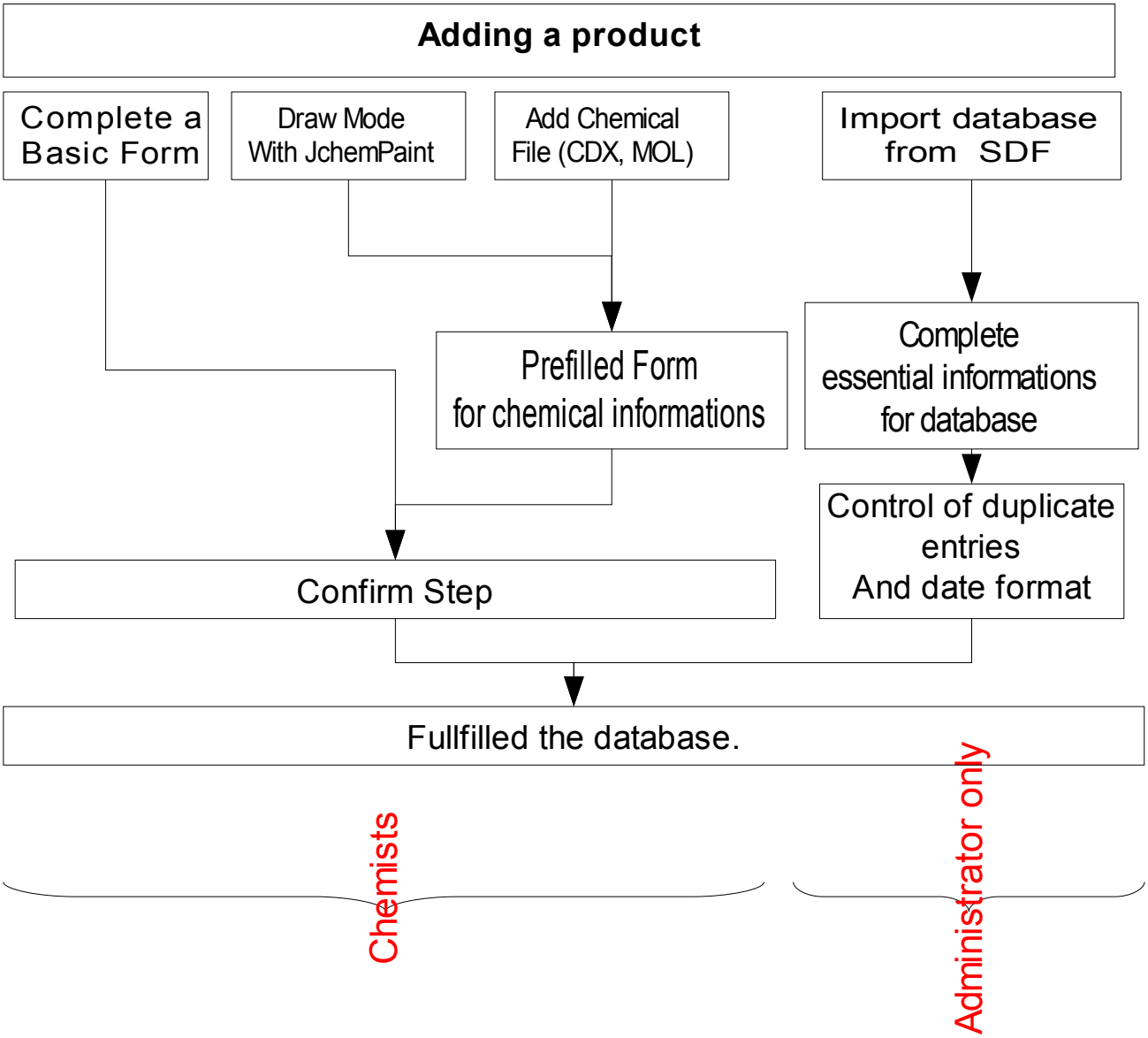
```
#vim /etc/php5/apache2/conf.d/suhosin.ini
```

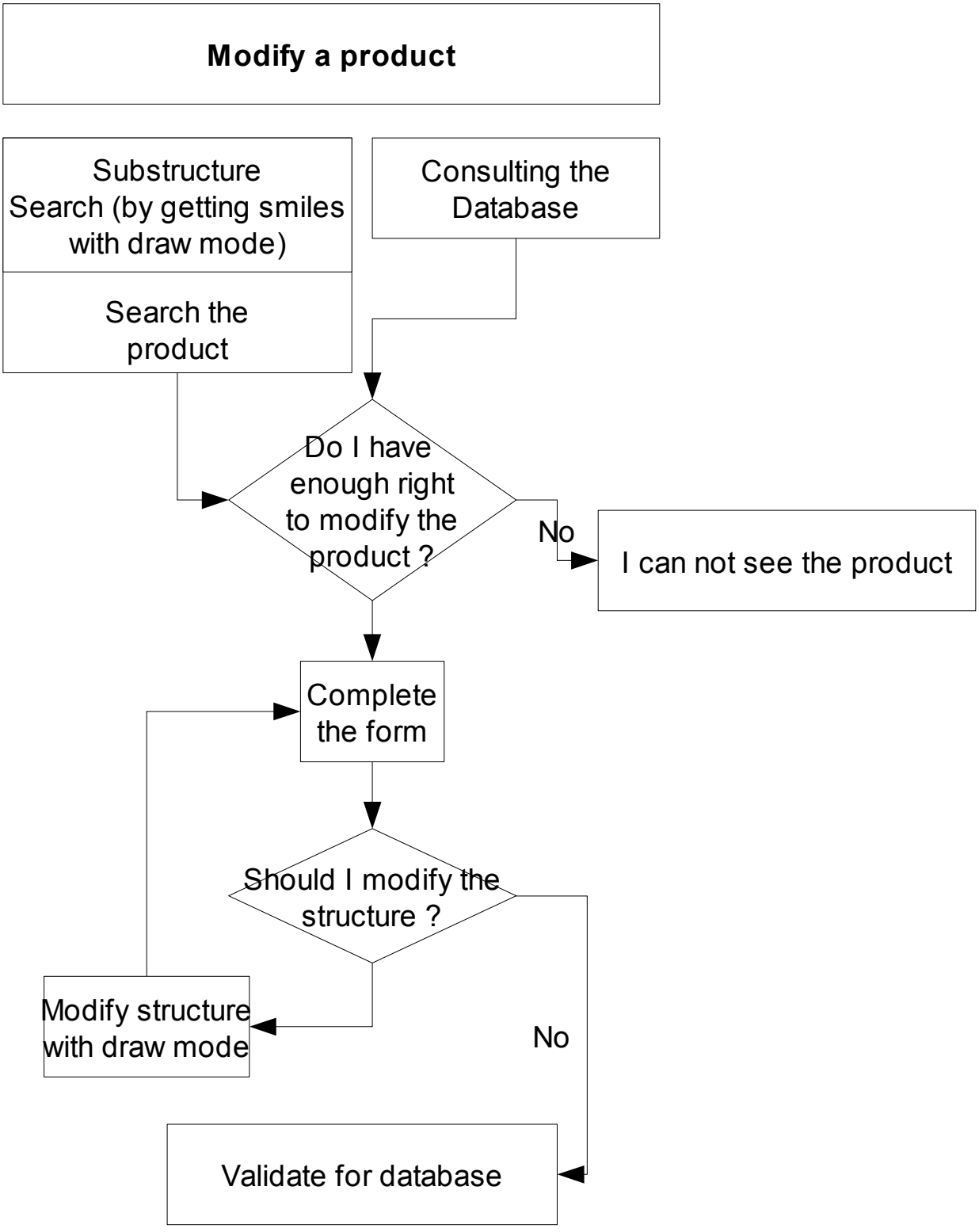
Uncomment and initialize the following values :

```
suhosin.post.max_totalname_length = 10000
suhosin.post.max_value_length = 65000
suhosin.post.max_vars = 10000
suhosin.request.max_totalname_length = 10000
suhosin.request.max_value_length = 65000
suhosin.request.max_vars = 10000
suhosin.cookie.max_totalname_length = 10000
suhosin.cookie.max_value_length = 10000
suhosin.cookie.max_vars = 10000
```

Then, *suhosin* should not be a problem...

General patterns and overall functionning




Modify a product
Substructure Search (by getting smiles with draw mode)
Consulting the Database
Search the product
Do I have enough right to modify the product ?
No
I can not see the product
Complete the form
Should I modify the structure ?
Modify structure with draw mode
No
Validate for database

	Add a product	**See *my* products**	**Search/modify *my* products**	**Delete *my* products**	**See other group members products**	**Modify other group members products**	**Delete other group members products**	**See/Modify all products**	**See/Modify/Delete all products, users and groups**
Visitor	Y	N	N	N	N	N	N	N	N
Temporary user	Y	Y	Y	N	N	N	N	N	N
Active user *No group*	Y	Y	Y	Y	N	N	N	N	N
Member of group "Read Only"	Y	Y	Y	Y	Y	N	N	N	N
Member of group "Read and modify"	Y	Y	Y	Y	Y	Y	N	N	N
Member of group "Read/Modify/ Delete"	Y	Y	Y	Y	Y	Y	Y	N	N
Member of group "Read/Modify all"	Y	Y	Y	Y	Y	Y	N	Y	N
Member of group "Read/Modify/ Delete all"	Y	Y	Y	Y	Y	Y	Y	Y	Y
Administrator	Y	Y	Y	Y	Y	Y	Y	Y	Y

Directory tree

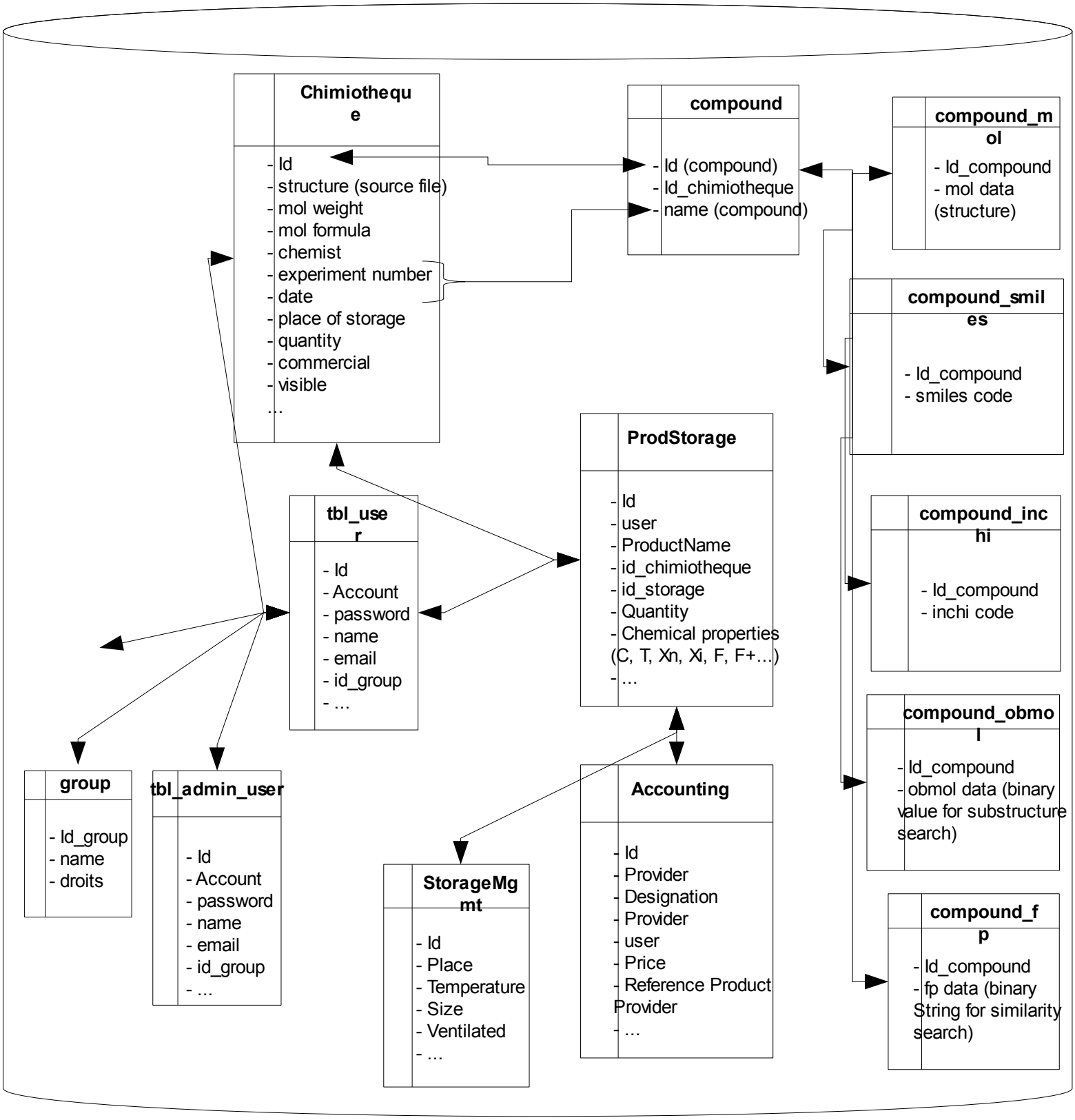
APACHE DOCUMENT ROOT
> tcpdf
> PHPExcel
> chemaztech
- ->> config
- ->> core
- ->> export
- ->> formulaires
- ->> fpdi
- ->> gestion
- ->> img
- ->> import
- ->> include
- ->> jchempaint
- ->> jmol
  - ->>> mol
  - ->>> ...
- ->> maintenance
- ->> pdf
- ->> zip
- ...

- MyChem Problems

---> http://mychem.sourceforge.net/

Notes and references

OpenSource : http://en.wikipedia.org/wiki/Open_source
ChemiSQL : http://chemdb.sourceforge.net/wiki/index.php/
MyChem : http://mychem.sourceforge.net/
OrChem : http://orchem.sourceforge.net/
pgchem : http://pgfoundry.org/projects/pgchem/
MDL : http://www.symyx.com/
ChemAxon : http://www.chemaxon.com/
libcairo : http://www.cairographics.org/
OpenBabel : http://openbabel.org/wiki/Main_Page
JchemPaint : http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint
CDK : http://sourceforge.net/apps/mediawiki/cdk/index.php?title=Main_Page
Jmol : http://jmol.sourceforge.net/
GNU GPL v2 license : http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
SMILES : http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html
OBMol object : http://openbabel.org/wiki/OBMol
MySQL : http://www.mysql.com/
TCPDF : http://www.tecnick.com/public/code/cp_dpage.php?aiocp_dp=tcpdf
FPDI : http://www.setasign.de/products/pdf-php-solutions/fpdi/
PHPEXcel : http://phpexcel.codeplex.com/

Appendix

Database pattern :