Instruction to Virtual IR Lab

(Manual for administrators)

Virtual IR Lab is a web toolkit developed by IR Lab in University of Delaware. It helps users to implement and test retrieval functions in a very convenient way. It should be useful for both education and research.

1. Install.

Virtual IR Lab consists of two parts: ***VirtualIRLab-p*** contains the files essential for the web service while ***Gougou*** is the codebase of the binary executable which handles the background ranking part.

Both of them can be installed on typical Linux environment. We will introduce how to install the toolkits step by step.

(1) Pre-requirement.

To install the toolkit, the following packages are needed:

a. LAMP (For *VirtualIRLab-p*)

b. C++ compiler (e.g. g++. For *Gougou*)

(2) Compile *Gougou*.

- Decompress gougou package:

```
tar –xf gougou.tar.gz
```

- Compile the toolkit:

```
cd gougou
make -f MakefileLibAll.app
make -f MakefileAll.app
```

If there is no error, you can move to next step.

(3) Setup the database.

In this step, you need to setup your mysql database and make it accessible through local web service through socket.

- Create an empty database:

```
mysql > create database virtualIRLab;
```

- Create a web user who has fully access to the database:

```
mysql > CREATE USER 'VirtualIRUser'@'localhost' IDENTIFIED BY 'Virtual';
mysql > GRANT ALL PRIVILEGES ON virtualIRLab.* TO
'VirtualIRUser'@'localhost';
```

(4) Install and configure the files for web service.

- Decompress V*irtualIRLab-p* package:

```
tar -xf VirtualIRLab-p.tar.gz
```

- Copy the directory to web service folder:

```
rsync –az VirtualIRLab-p /var/www/html/
```

- Copy the binary executable files in *gougou* to the web service folder:

```
cp gougou/bin/* /var/www/html/VirtualIRLab-p/source/
```

- And make sure they are executable:

```
chmod a+x /var/www/html/VirtualIRLab-p/source/BuildIndexCmd
chmod a+x /var/www/html/VirtualIRLab-p/source/getSnippet
chmod a+x /var/www/html/VirtualIRLab-p/source/retrieval-flexible-web
chmod a+x /var/www/html/VirtualIRLab-p/source/retrieval-f-qry
chmod a+x /var/www/html/VirtualIRLab-p/source/showDoc
```

- Modify the conf/conf.php to fit your configurations.

```
<?php
  define('UploadPath','upload/');
  define('BuildIndexProgram','source/BuildIndexCmd');
  define('IndexPath','index/');
  define('MysqlHost','localhost');
  define('MysqlUser','VirtualIRLabUser');
  define('MysqlPassword','Virtual');
  define('MysqlPort',3666);
  define('MysqlDatabase','virtualIRLab');
  define('UserDataPath','users/');
  define('RetrievalProgramDirect','source/retrieval-flexible-web');
?>
~
```

You need to change “MysqlUser”, ”MysqlPassword”, ”MysqlPort”, ”MysqlDatabase” to match your setup in the previous section. Save and exit the file.

(5) Initial the system.

- Open any browser and initial the system by typing and confirming the following command:

```
http://localhost/VirtualIRLab-p/initial.php
```

Don’t run it twice and it may lead problems!

If you got problems or your run initial.php multiple times, please empty the database and run initial.php again.

- Disable or delete *initial.php* for security purpose by doing either of the following:

```
chmod 000 initial.php
rm initial.php
```

Now the system is ready to use!

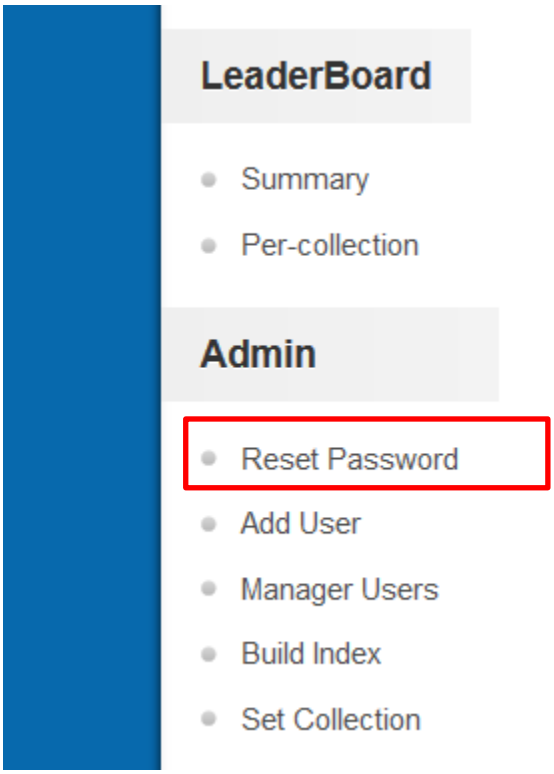
(6) First login:

Use browser to open the login page and login with

Username: **admin**

Password: **admin**

After you login, please reset password for security purpose.



2. Administration

In this section, we would like to tell you how to manage the system as administrator.

(1) Add/Manage users.

A user who registers the system must wait for your approve to use the system. You can approve or decline the requirement at the “add user” menu. You can also add user here without registration here.

You can also manage the users in the “manage user” menu. It will show the statistics of the users. You can also delete users or reset their password to the same as their username.

loginName	type	first name	last name	affiliation	email	functionGroup #	function #	engine #	control
admin	255					0	0	0	resetPasswd
tt1	1					0	0	0	resetPasswd Delete

(2) Add/Manage Collections.

To play with the toolkit, you need to upload at least one data collections.

Firstly you need to build an index in the “Build Index” menu. The index is built based on a text file in TREC format. The content of the file should look like:

```
<DOC>
<DOCNO>Doc1 </DOCNO>
<TEXT>
…
…
</TEXT>
</DOC>
<DOC>
<DOCNO>Doc2</DOCNO>
<TEXT>
…
</TEXT>
</DOC>
```

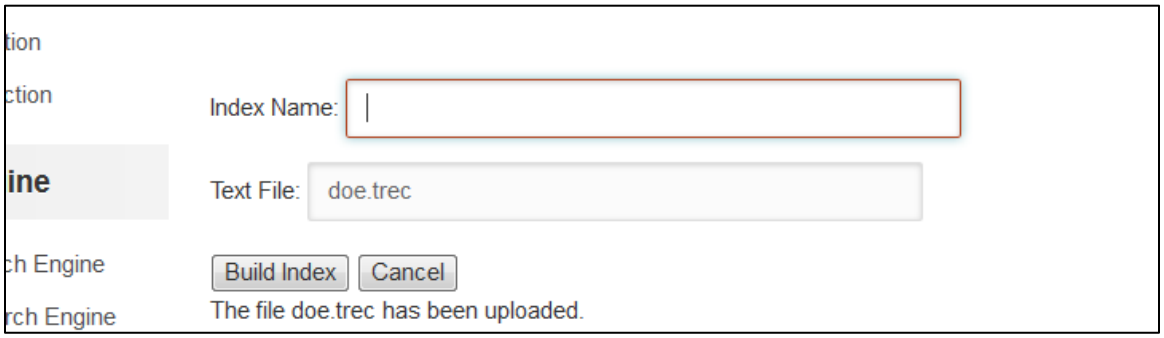
For large file, it may take long time to upload it (e.g. several mins or tens of mins). If you upload failed, you may need to change the php setting in “`/etc/php5/apache2/php.ini`” (if you are using Ubuntu) in order to upload large files. You may need to change the following lines in php setting and restart the web server.

```
; Maximum amount of memory a script may consume (128MB)
; http://php.net/memory-limit
memory_limit = 8096M
```

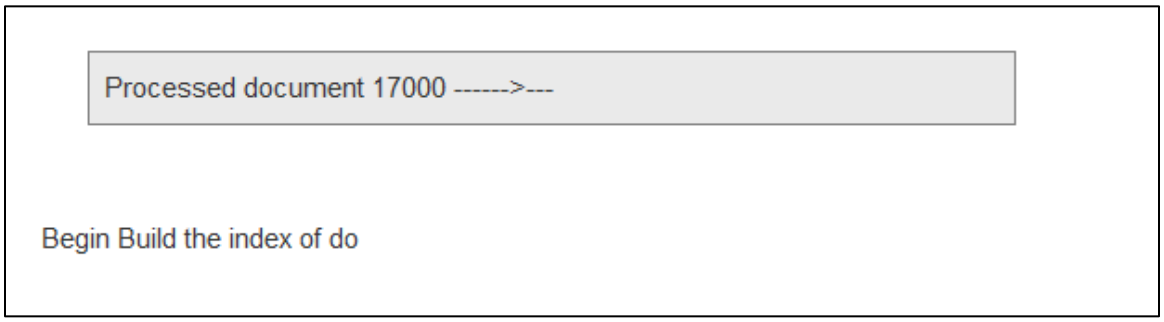
```
; Maximum allowed size for uploaded files.
; http://php.net/upload-max-filesize
upload_max_filesize = 8096M
```

```
; Maximum size of POST data that PHP will accept.
; Its value may be 0 to disable the limit. It is ignored if POST data reading
; is disabled through enable_post_data_reading.
; http://php.net/post-max-size
post_max_size = 8096M
```

After you upload the file, you can give a name and begin building the index.

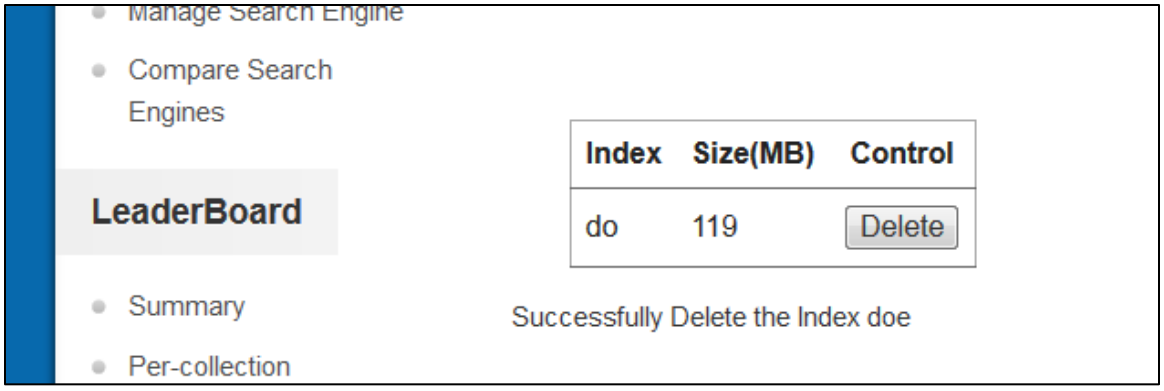


It will show the process when building the index

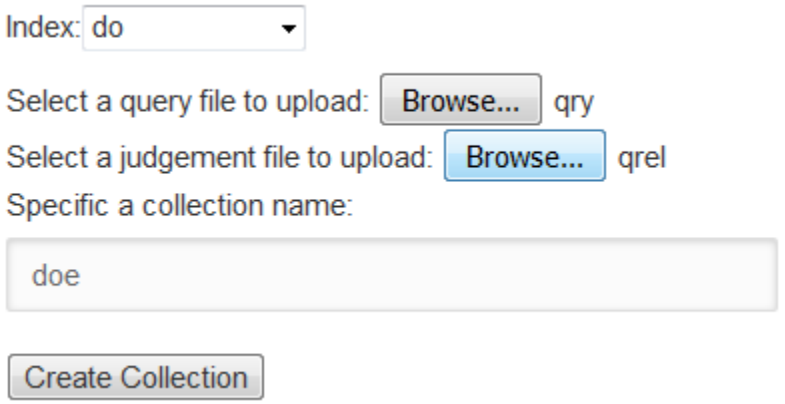


At the end of process, the image may freeze but the backend process is still running. So please be patient.

After building the index, you can see the statistics of the index and delete it as well.



The next step is to add queries and judgement associated with the index in the menu of “Set Collection”.



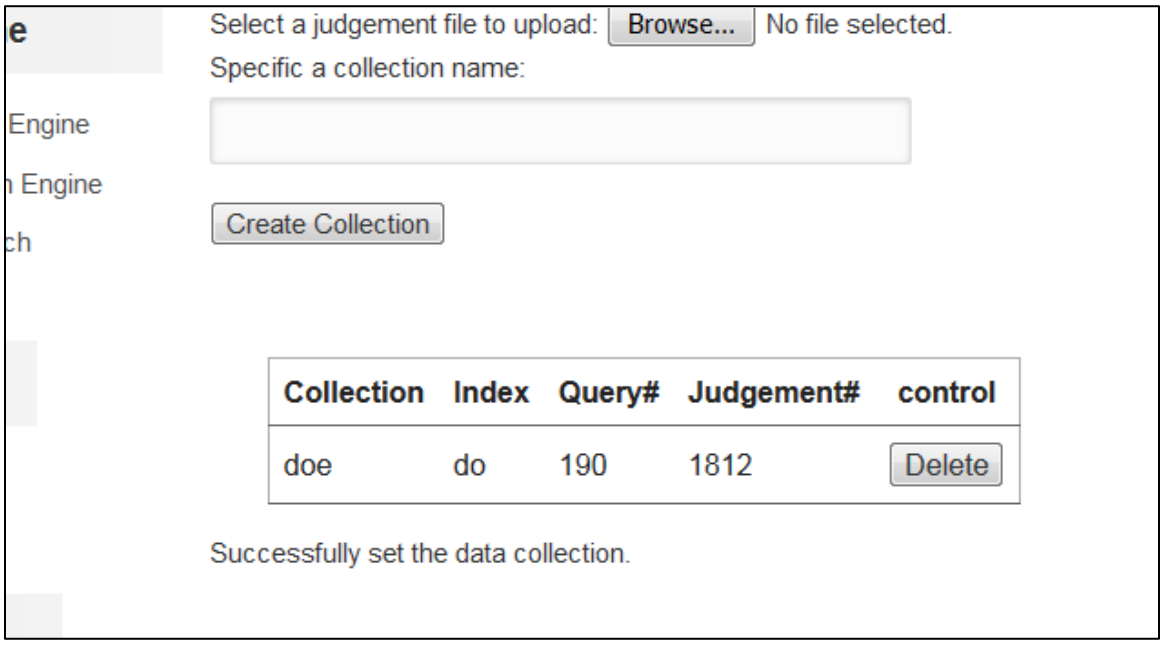
Query file should follow the TREC format:

```
11: space program

12: water pollution

…
```

And the judgement file should follow TREC format as well:

```
102   DOC3   1

102   DOC33   1

11    DOC1   1

11    DOC123   1

…
```

Name the collection and create it. It may take several minutes to upload the file and insert into database as well.
You should be able to see the statistics of the data collection as well.



It should be ready to use then☺