

Anexo I

Manual de instalación

Webbot para Datos Bibliométricos (WB).

Integrantes :

De Domingo Gonzalo, Guastadisegni Federico, Rodríguez Nahuel.

Cátedra :

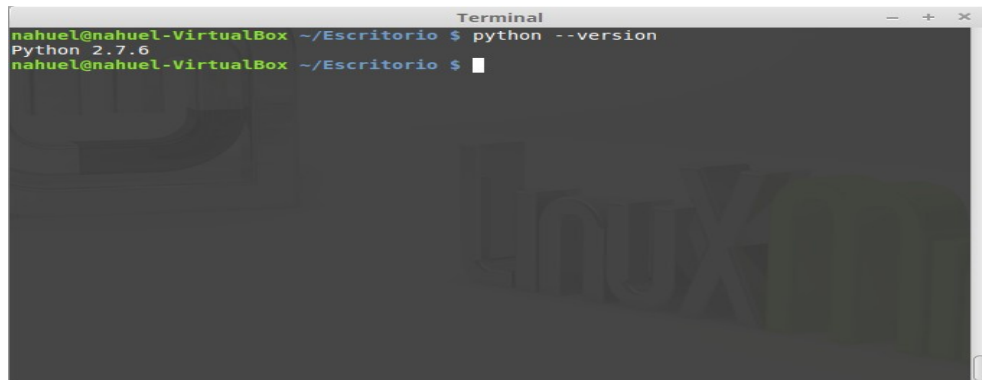
Proyecto de Software
Licenciatura en Sistemas.

2015

Requisitos

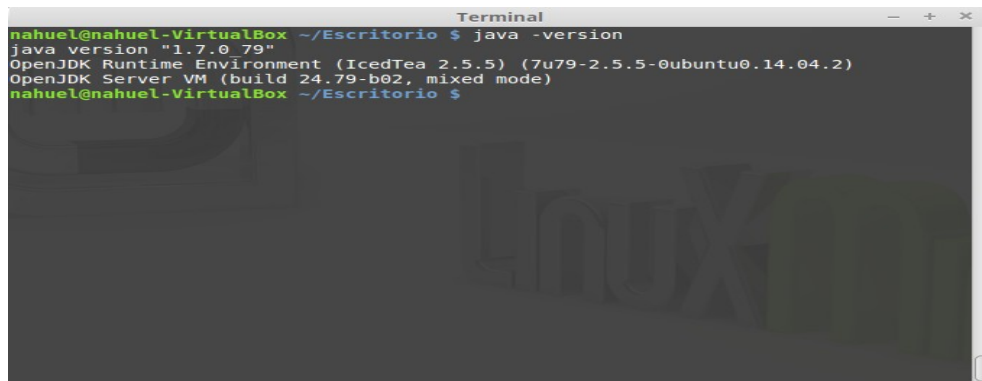
Se contemplan como pre-requisitos para la instalación, tener configurado:

- Python versión 2.7.x.

A terminal window titled "Terminal" with a dark background. The prompt is "nahuel@nahuel-VirtualBox ~/Escritorio". The command "python --version" has been entered, and the output is "Python 2.7.6".

```
nahuel@nahuel-VirtualBox ~/Escritorio $ python --version
Python 2.7.6
nahuel@nahuel-VirtualBox ~/Escritorio $
```

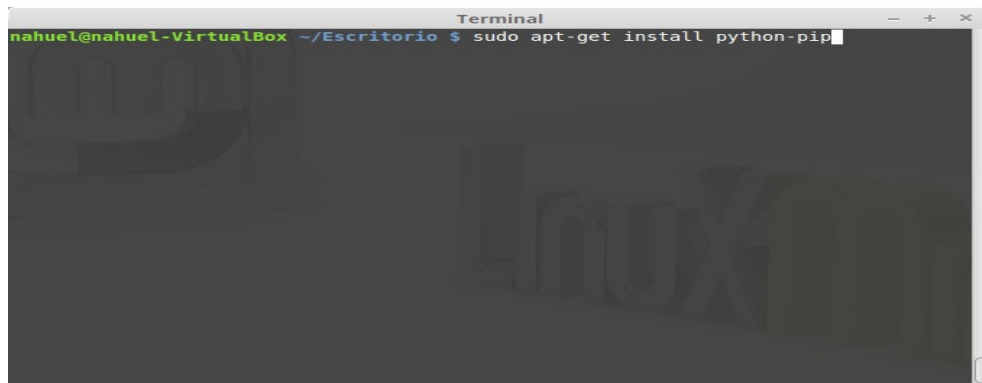
- Java versión 1.7.x.

A terminal window titled "Terminal" with a dark background. The prompt is "nahuel@nahuel-VirtualBox ~/Escritorio". The command "java -version" has been entered, and the output shows the Java version and runtime environment details.

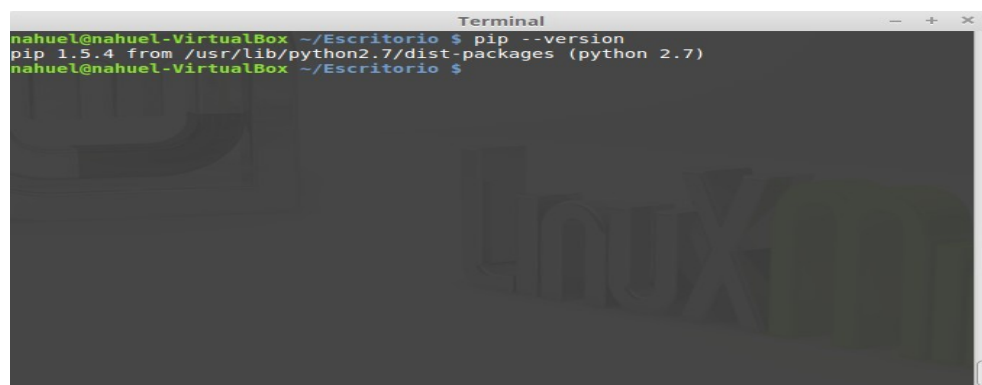
```
nahuel@nahuel-VirtualBox ~/Escritorio $ java -version
java version "1.7.0_79"
OpenJDK Runtime Environment (IcedTea 2.5.5) (7u79-2.5.5-0ubuntu0.14.04.2)
OpenJDK Server VM (build 24.79-b02, mixed mode)
nahuel@nahuel-VirtualBox ~/Escritorio $
```

Instrucciones

1. Instalar PIP 1.5.4 con el siguiente comando:
 - Sudo apt-get install python-pip.

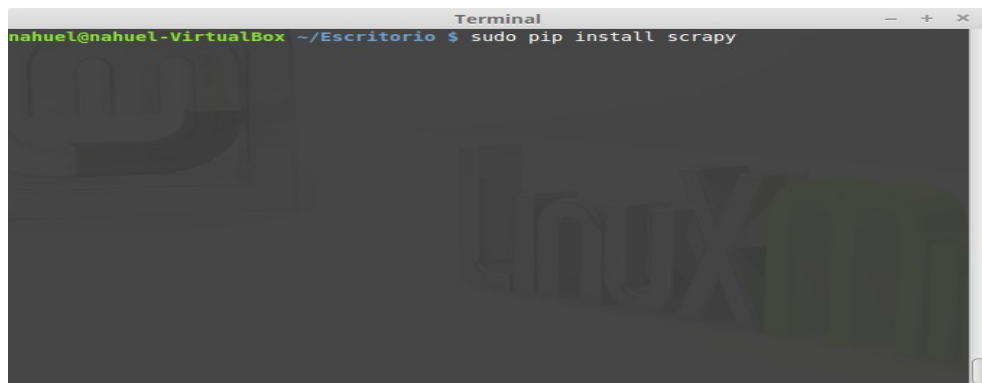


```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo apt-get install python-pip
```

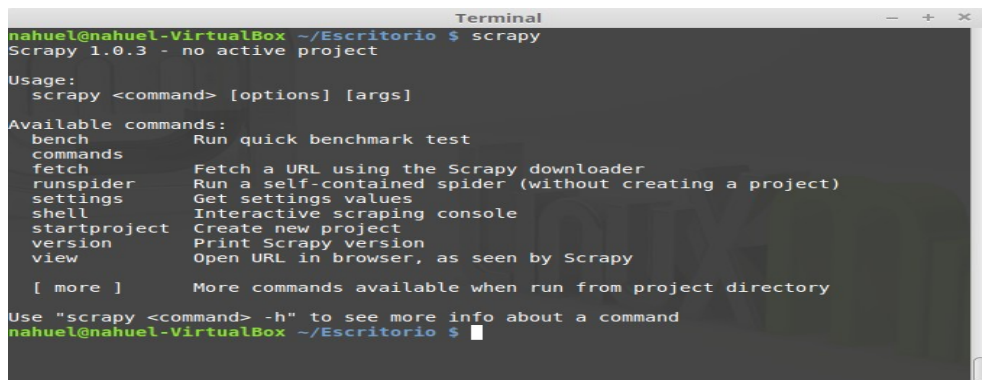


```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ pip --version
pip 1.5.4 from /usr/lib/python2.7/dist-packages (python 2.7)
nahuel@nahuel-VirtualBox ~/Escritorio $
```

2. Instalar Scrapy 1.0.3 con el siguiente comando:
 - Sudo pip install scrapy.



```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo pip install scrapy
```



```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ scrapy
Scrapy 1.0.3 - no active project

Usage:
  scrapy <command> [options] [args]

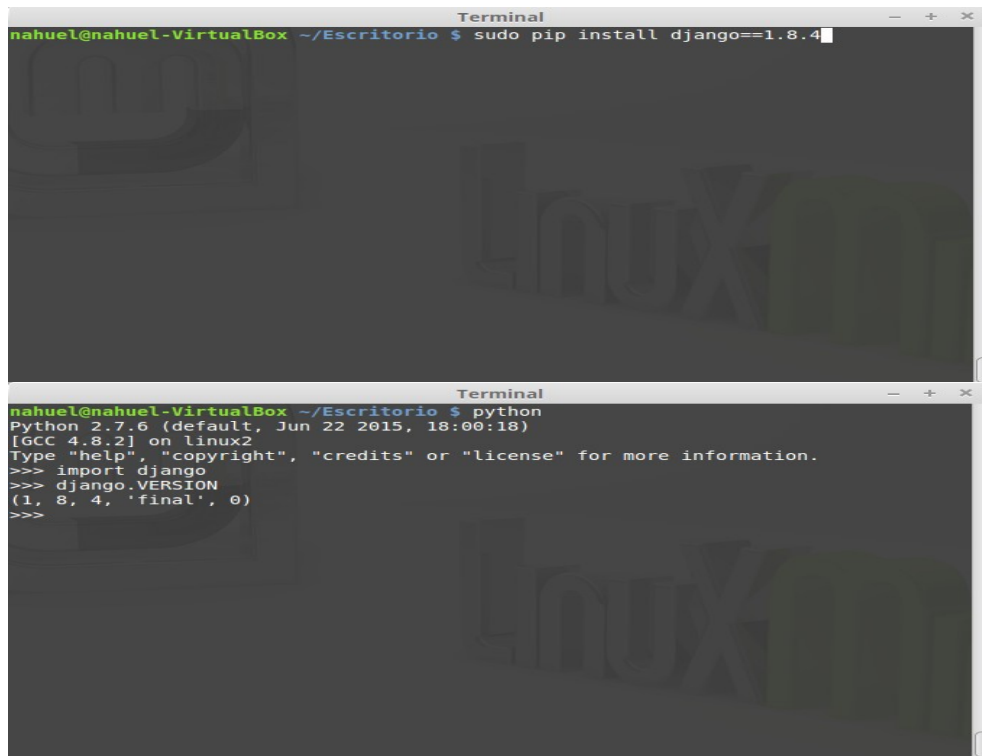
Available commands:
bench          Run quick benchmark test
commands       Run a self-contained spider (without creating a project)
fetch          Fetch a URL using the Scrapy downloader
runspider      Run a self-contained spider (without creating a project)
settings       Get settings values
shell          Interactive scraping console
startproject   Create new project
version        Print Scrapy version
view           Open URL in browser, as seen by Scrapy

[ more ]       More commands available when run from project directory

Use "scrapy <command> -h" to see more info about a command
nahuel@nahuel-VirtualBox ~/Escritorio $
```

3. Instalar Django 1.8.4 con el siguiente comando:

- Sudo pip install Django==1.8.4.



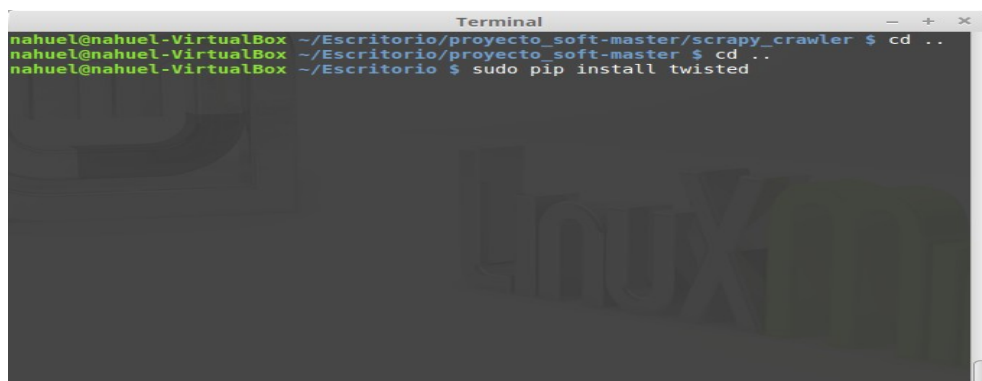
The first terminal window shows the command `sudo pip install django==1.8.4` being executed. The second terminal window shows the command `python` being executed, followed by `>>> import django`, `>>> django.VERSION`, and `>>>`, resulting in the output `(1, 8, 4, 'final', 0)`.

```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo pip install django==1.8.4

Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ python
Python 2.7.6 (default, Jun 22 2015, 18:00:18)
[GCC 4.8.2] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import django
>>> django.VERSION
(1, 8, 4, 'final', 0)
>>>
```

4. Instalar Twisted con el siguiente comando:

- Sudo pip install twisted.

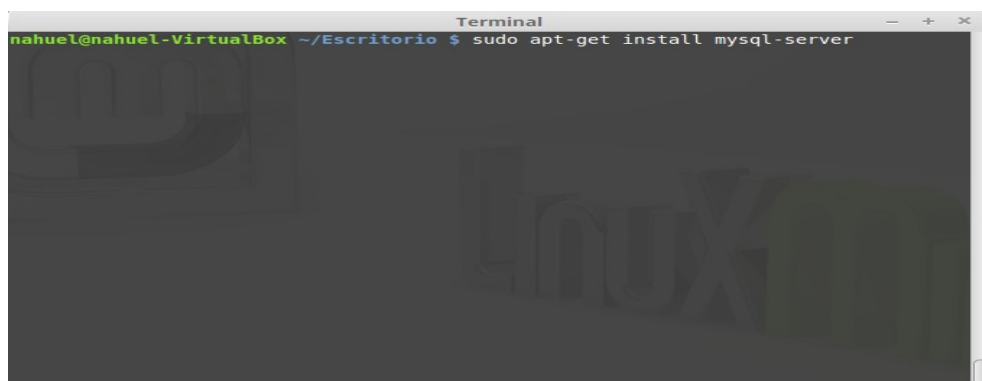


The terminal window shows the command `sudo pip install twisted` being executed.

```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio/proyecto_soft-master/scrappy_crawler $ cd ..
nahuel@nahuel-VirtualBox ~/Escritorio/proyecto_soft-master $ cd ..
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo pip install twisted
```

5. Instalar MySQL con el siguiente comando:

- Sudo apt-get install mysql-server.

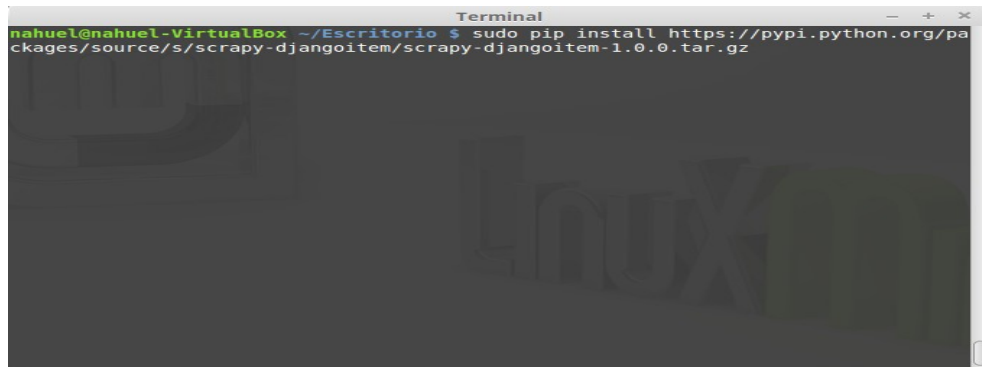


The terminal window shows the command `sudo apt-get install mysql-server` being executed.

```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo apt-get install mysql-server
```

6. Instalar DjangoItem con el siguiente comando:

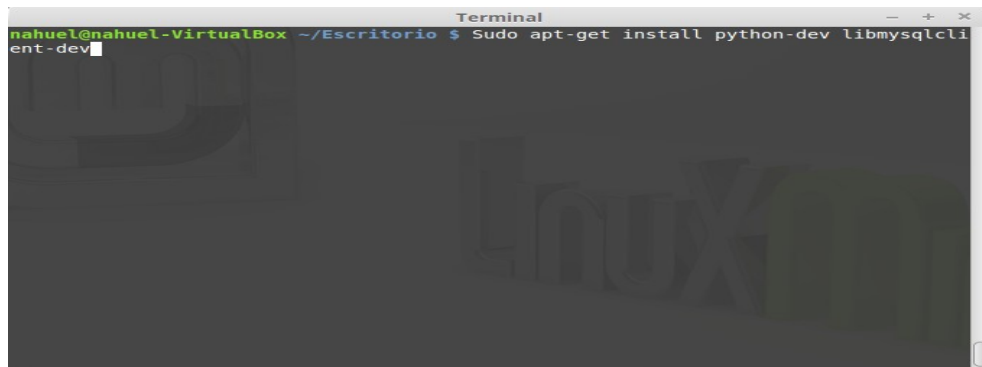
- Sudo pip install <https://pypi.python.org/packages/source/s/scrappy-djangoitem/scrappy-djangoitem-1.0.0.tar.gz>



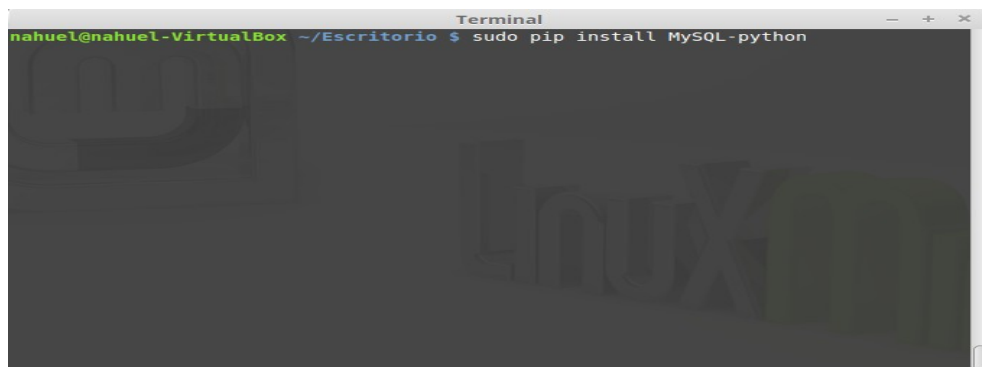
```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo pip install https://pypi.python.org/packages/source/s/scrappy-djangoitem/scrappy-djangoitem-1.0.0.tar.gz
```

7. Una vez instalado DjangoItem, se requiere un conector para la base de datos, por lo que se deben ejecutar los siguientes comandos:

- Sudo apt-get install python-dev libmysqlclient-dev.
- Sudo pip install MySQL-python.



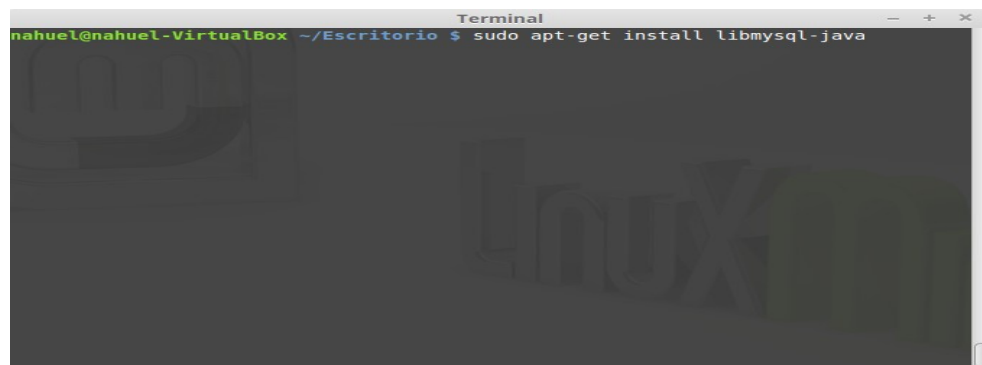
```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ Sudo apt-get install python-dev libmysqlclient-dev
```



```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo pip install MySQL-python
```

8. Instalar el conector Java para Solr con el siguiente comando:

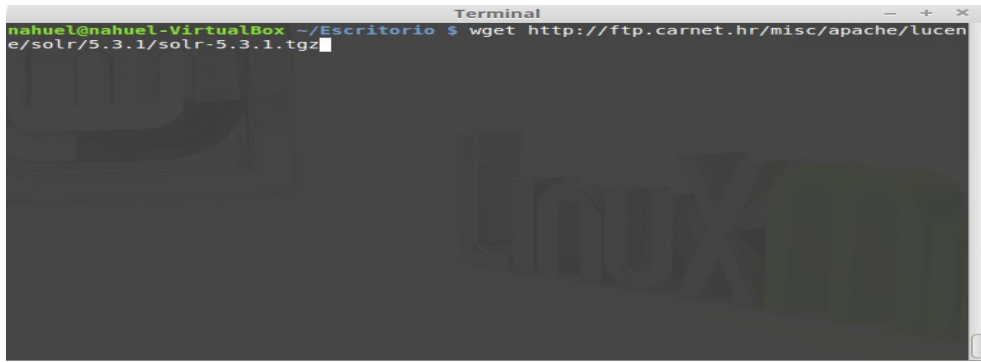
- Sudo apt-get install libmysql-java.



```
Terminal
nahuel@nahuel-VirtualBox ~/Escritorio $ sudo apt-get install libmysql-java
```

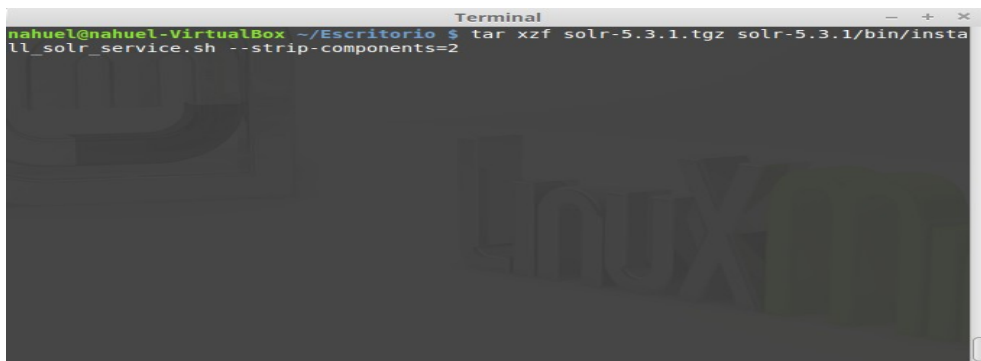
9. Descargar el archivo comprimido de Apache Solr con el siguiente comando:

- Wget <http://ftp.carnet.hr/misc/apache/lucene/solr/5.3.1/solr-5.3.1.tgz>



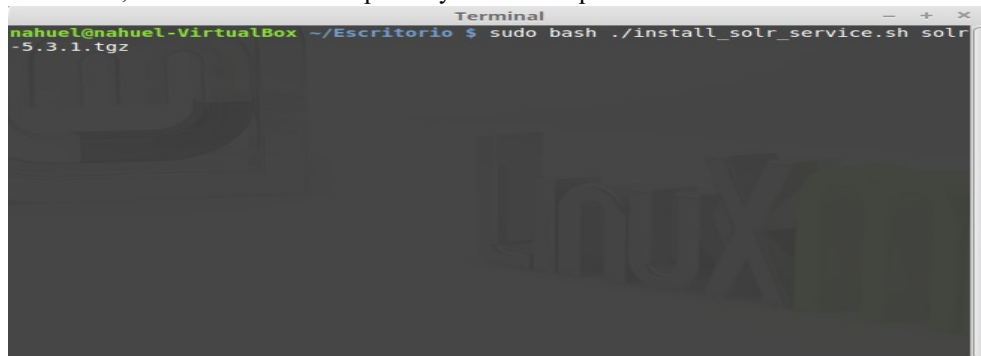
10. Una vez descargado, hacer un CD hasta el directorio donde esté el paquete, y ejecutar el siguiente comando:

- Tar xzf solr-5.3.1.tgz solr-5.3.1/bin/install_solr_service.sh --strip-components=2.



11. Posteriormente, ejecutar el siguiente comando:

- Sudo bash ./install_solr_service.sh solr-5.3.1.tgz.
Por defecto, Solr se instalará en /opt/solr y utilizará el puerto 8983.



12. Una vez instalado Solr, bajar el siguiente conector:

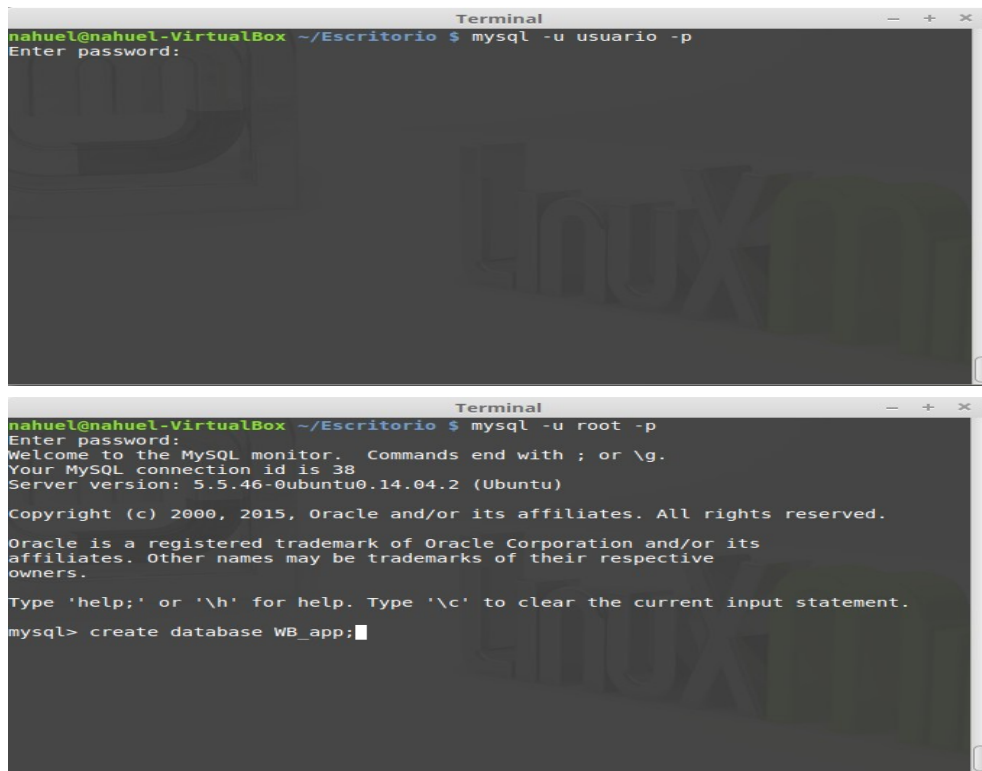
- <http://cdn.mysql.com/archives/mysql-connector-java-5.1/mysql-connector-java-5.1.32.tar.gz>

Ponerlo en /opt/solr-5.3.1/contrib/dataimporthandler (se debe especificar esta ruta en el solrconfig.xml de cada core que se utilice).



13. Ejecutar los siguientes comandos en la consola:

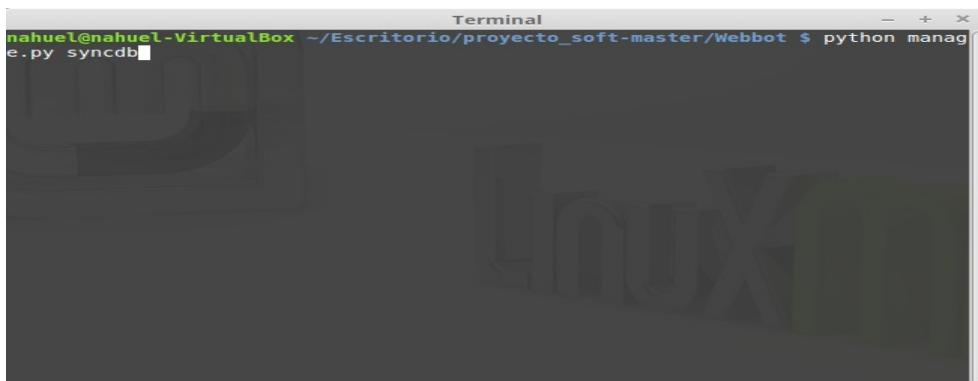
- `mysql -u usuario -p` (donde usuario es el usuario administrador de las bases de datos). Presionar enter, ingresar contraseña de administrador y presionar nuevamente enter.
- `create database WB_app;`



The first terminal screenshot shows the command `mysql -u usuario -p` being entered, followed by the prompt "Enter password:". The second terminal screenshot shows the command `mysql -u root -p` being entered, followed by the MySQL welcome message and the command `create database WB_app;` being entered at the `mysql>` prompt.

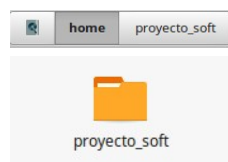
14. Ir a `/proyecto_soft/Webbot` y ejecutar por consola el siguiente comando:

- `python manage.py syncdb`.



The terminal screenshot shows the command `python manage.py syncdb` being entered in the directory `~/Escritorio/proyecto_soft-master/Webbot`.

15. Para ejecutar el Webbot la carpeta “proyecto_soft” debe encontrarse en el directorio `/home`, como se muestra a continuación, para el correcto funcionamiento del Script:



Dentro de esta carpeta, encontrará un archivo llamado “Webbot.sh”:



Para poder ejecutarlo, primero hay que abrir la terminal, posicionarnos en la carpeta “proyecto_soft”, y ejecutar el siguiente comando:

- `chmod +x Webbot.sh`. Este comando hará ejecutable el archivo.

Luego de hacer ejecutable el archivo, simplemente haga doble click sobre el mismo. Se abrirá la siguiente ventana:



Haga click en el botón “Ejecutar en un terminal” y espere a que termine la ejecución del Webbot. Este proceso puede demorar unos minutos.

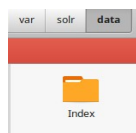
```
Terminal
2015-11-05 21:17:13 [scrapy] INFO: Scrapy 1.0.3 started (bot: scrapy_crawler)
2015-11-05 21:17:13 [scrapy] INFO: Optional features available: ssl, http11
2015-11-05 21:17:13 [scrapy] INFO: Overridden settings: {'NEWSPIDER_MODULE': 'scrapy_crawler.spiders', 'SPIDER_MODULES': ['scrapy_crawler.spiders'], 'BOT_NAME': 'scrapy_crawler'}
2015-11-05 21:17:13 [scrapy] INFO: Enabled extensions: CloseSpider, TelnetConsole, LogStats, CoreStats, SpiderState
2015-11-05 21:17:13 [scrapy] INFO: Enabled downloader middlewares: HttpAuthMiddleware, DownloadTimeoutMiddleware, UserAgentMiddleware, RetryMiddleware, DefaultHeadersMiddleware, MetaRefreshMiddleware, HttpCompressionMiddleware, RedirectMiddleware, CookiesMiddleware, ChunkedTransferMiddleware, DownloaderStats
2015-11-05 21:17:13 [scrapy] INFO: Enabled spider middlewares: HttpErrorMiddleware, OffsiteMiddleware, RefererMiddleware, UrlLengthMiddleware, DepthMiddleware
2015-11-05 21:17:13 [scrapy] INFO: Enabled item pipelines: ScrapAutores
2015-11-05 21:17:13 [scrapy] INFO: Spider opened
2015-11-05 21:17:13 [scrapy] INFO: Crawled 0 pages (at 0 pages/min), scraped 0 items (at 0 items/min)
2015-11-05 21:17:13 [scrapy] DEBUG: Telnet console listening on 127.0.0.1:6023
```

Una vez que el proceso concluya, se cerrará la ventana de la terminal. Eso quiere decir que los datos procesados ya están almacenados en la base de datos.

16. Preparación del core de Solr : En la carpeta “proyecto_soft”, notará que existe una carpeta que se llama Solr. Dentro de la misma, se encuentra la carpeta Index.



Deberá copiar esa carpeta, y pegarla en /var/solr/data:



En el caso de que no se lo permita, por cuestiones de permisos, debe hacer click derecho en un lugar vacío de la carpeta y hacer click en “Abrir con permisos administrativos”. Se le pedirá su contraseña, y se abrirá la carpeta con privilegios elevados:



Lo que debe hacer ahora, es resetear el servicio de Solr, para que este reconozca el core. Para ello, vamos a la terminal y ejecutamos el comando:

- `sudo service solr restart`.

Esperamos a que termine y el core estará listo para ser utilizado.

Aclaración : En el caso de haber creado la base de datos con un usuario/contraseña que no sea root/root, deberá cambiar estos datos en /Index/conf/db-data-config.xml:

```
<dataSource type="JdbcDataSource"
  driver="com.mysql.jdbc.Driver"
  url="jdbc:mysql://localhost:3306/WB_app"
  user="root"
  password="root" />
```