Deepdive 环境搭建及操作教程

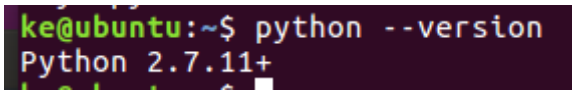
Deepdive 是由斯坦福大学 InfoLab 实验室开发的一个开源知识抽取系统。它通过弱监督学习，从非结构化的文本中抽取结构化的关系数据 。CNDeepdive 修改了自然语言处理的model包，使它支持中文，并提供中文tutorial。

一、虚拟机环境配置

虚拟机安装教程：https://blog.csdn.net/weixin_74195551/article/details/127288338

推荐安装 ubuntu 16.04 版本

查看 python 版本

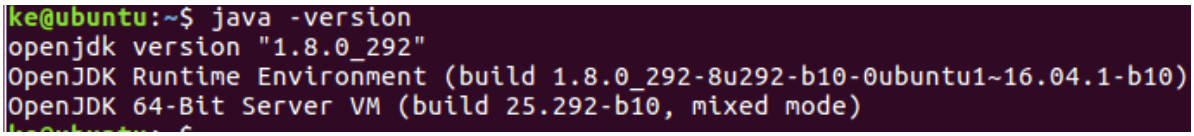


安装 openjdk-8-jdk

```
sudo apt-get install -y openjdk-8-jdk
```

查看 java 版本，验证是否安装成功

```
java -version
```

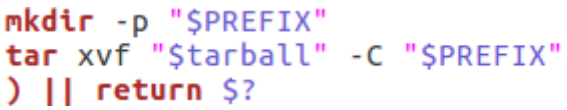


二、安装 CNDeepdive

下载 CNdeepdive （https://github.com/emperinter/CNdeepdive/tree/master/CNdeepdive）后，进入 CNdeepdive 目录下，为 install.sh 和 nlp_setup.sh 文件增加可执行权限

```
chmod +x install.sh
chmod +x nlp_setup.sh
```

打开 install.sh 文件，修改第193行内容



运行 install.sh 文件，输入 1 安装 deepdive

```
ke@ubuntu:~/CNdeepdive$ ./install.sh
### DeepDive installer for Ubuntu
1) deepdive                   6) postgres
2) deepdive_docker_sandbox    7) postgres_xl
3) deepdive_example_notebook  8) run_deepdive_tests
4) deepdive_from_release      9) spouse_example
5) jupyter_notebook
# Install what (enter to repeat options, a to see all, q to quit, or a number)?
1
## Starting installation for deepdive
## Starting installation for deepdive_from_release
# Installing DeepDive release v0.8-STABLE...
# Press any key to select a different release or Enter to proceed...
```

出现两个 finished 证明安装成功

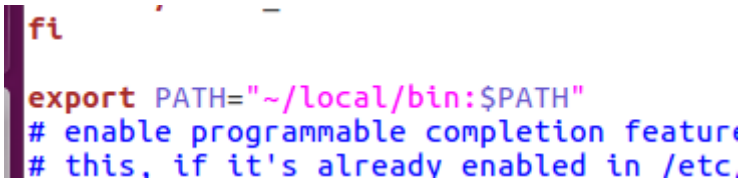
```
+ sudo localedef -i en_US -f UTF-8 en_US.UTF-8
## Finished installation for _deepdive_runtime_deps
## Finished installation for deepdive
# Install what (enter to repeat options, a to see all, q to quit, or a nu
q
```

完成后需要配置CNdeepdive的环境变量

```
gedit ~/.bashrc
```

添加下面的语句

```
export PATH="~/local/bin:$PATH"
```



运行下面的语句让改动生效

```
source ~/.bashrc
```

在控制台输入 deepdive，出现下面的输出证明配置成功

```
ke@ubuntu:~/CNdeepdive$ deepdive
# DeepDive (v0.8.0-79-g28a58de)
# Usage: deepdive COMMAND [ARG]...

# To enable bash completion for deepdive commands, run:
source $(deepdive whereis installed etc/deepdive_bash_completion.sh)

No COMMAND given

# Available COMMANDs are:
deepdive  check    #    Checks errors in compiled DeepDive app
deepdive  compile  #    Compiles DeepDive source code into executables under run/
deepdive  compute  #    Runs a UDF using a computer against the database
deepdive  create   #    Creates a table/view in the database
deepdive  db       #    Exposes lower level database operations provided by the driver
deepdive  do       #    Runs necessary processes to get something done
```

三、安装 postgresql

运行 install.sh 文件，输入 6 安装 postgresql

```
ke@ubuntu:~/CNdeepdive$ ./install.sh
### DeepDive installer for Ubuntu
1) deepdive                   6) postgres
2) deepdive_docker_sandbox    7) postgres_xl
3) deepdive_example_notebook  8) run_deepdive_tests
4) deepdive_from_release      9) spouse_example
5) jupyter_notebook
# Install what (enter to repeat options, a to see all, q to quit, or a number)? 6
## Starting installation for postgres
```

出现 finished 证明安装成功

```
+ sudo cat /etc/postgresql/9.5/main/pg_hba.conf
+ sudo tee /etc/postgresql/9.5/main/pg_hba.conf
+ sudo service postgresql restart
++ rm -f /tmp/pg_hba.conf.uCjrD2d
## Finished installation for postgres
# Install what (enter to repeat options, a to see all, q to quit, or a number)?
```

```
ke@ubuntu:~/CNdeepdive$ sudo su - postgres
[sudo] password for ke:
postgres@ubuntu:~$
```

四、安装 nlp

运行 nlp_setup.sh 文件

```
ke@ubuntu:~/CNdeepdive$ ./nlp_setup.sh
Install Dependency.
Denpendency Already Installed.
```

五、Smoke Example

这个简单的例子是基于一个经典的马尔可夫逻辑网络例子来展示 deepdive 的概率推理功能。目的是用A=>B来推断一个人是否吸烟，以及一个人是否有可能患癌症。

推理规则:

1.如果A吸烟，那么A可能患有癌症。
2 .如果A和B是朋友，A吸烟，那么B也可能吸烟。

这些规则写在app.ddlog 中。

进入 postgresql 控制台

```
sudo su - postgres
psql
```

```
ke@ubuntu:~/CNdeepdive$ sudo su - postgres
[sudo] password for ke:
postgres@ubuntu:~$ psql
psql (9.5.25)
Type "help" for help.

postgres=#
```

创建用户 CREATE USER ke WITH PASSWORD '1222';

```
CREATE USER ke WITH PASSWORD 'your_password';
```

设置用户权限

```
ALTER ROLE ke CREATEROLE CREATEDB;
```

创建 smoke_test 数据库，并指定所有者

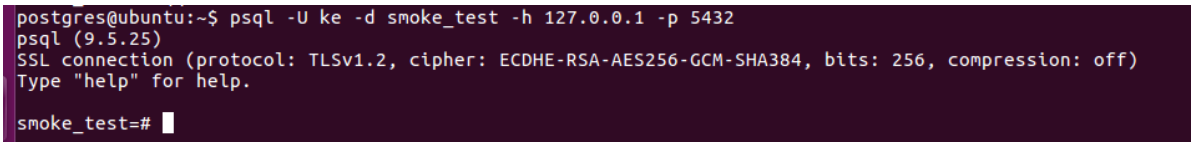
```
CREATE DATABASE smoke_test OWNER ke;
```

退出 postgresql 控制台

```
\q
```

查看是否创建成功

```
psql -U ke -d smoke_test -h 127.0.0.1 -p 5432
```



完成后退出 postgresql 控制台

进入 smoke 文件夹后，修改用户权限

```
chmod -R 777 ./
```

建立数据库配置文件

```
echo "postgresql://ke@localhost:5432/smoke_test" >db.url
```

编译

```
deepdive compile
```

运行

```
deepdive run
```

运行过程中会进入 vi，通过 :wq 退出即可

运行完成后查看结果

```
psql smoke_test
```

```
smoke_test=# \d
                              List of relations
 Schema |                          Name                           | Type  | Owner
--------+---------------------------------------------------------+-------+-------
 public | dd_factors_inf_imply_person_smokes_person_has_cancer    | table | ke
 public | dd_factors_inf_imply_person_smokes_person_smokes        | table | ke
 public | dd_graph_variables_holdout                              | table | ke
 public | dd_graph_variables_observation                          | table | ke
 public | dd_graph_weights                                        | view  | ke
 public | dd_inference_result_variables                           | table | ke
 public | dd_weights_inf_imply_person_smokes_person_has_cancer    | table | ke
 public | dd_weights_inf_imply_person_smokes_person_smokes        | table | ke
 public | friends                                                 | table | ke
 public | person                                                  | table | ke
 public | person_has_cancer                                       | table | ke
 public | person_has_cancer_label_calibration                     | view  | ke
 public | person_has_cancer_label_inference                       | view  | ke
 public | person_smokes                                           | table | ke
 public | person_smokes_label_calibration                         | view  | ke
 public | person_smokes_label_inference                           | view  | ke
(16 rows)
```

查看推理结果

```
select * from person_has_cancer_label_inference
```

```
smoke_test=# select * from person_has_cancer_label_inference;
 person_id | id | label | category | expectation
-----------+----+-------+----------+-------------
         1 |  0 |       |        1 |       0.624
         3 |  2 |       |        1 |       0.618
         5 |  4 |       |        1 |        0.58
         2 |  1 |       |        1 |       0.574
         4 |  3 |       |        1 |       0.573
         6 |  5 |       |        1 |       0.552
(6 rows)
```