

实践二 DeepDive关系抽取

课程名称： 知识工程 实验日期： 2023-6-18

班级： 网安一班 姓名： 贾庆林 学号： 3020244250

一、实践要求

1、使用DeepDive进行感兴趣领域知识抽取，需包括实体抽取和关系抽取。抽取的知识以三元组形式展示，要求总数量不少于30条三元组。

二、实践内容

- 1、DeepDive安装及环境配置
- 2、实体抽取
- 3、关系抽取

三、实践步骤（可包括核心代码与过程截图）

实验指导中给到的操作指令：

进入PostgreSQL控制台

```
sudo su - postgres  
psql
```

创建与系统用户同名的数据库用户并设密码（如果系统用户只有root，需要先在系统中创建一个系统用户：
sudo adduser user）

```
CREATE USER user WITH PASSWORD 'password';
```

设置用户权限

```
postgres=# ALTER ROLE user CREATEROLE CREATEDB;      或  
ALTER ROLE user CREATEROLE SUPERUSER;
```

创建用户数据库，这里为smoke_example，并指定所有者为刚才创建的用户

```
CREATE DATABASE smoke_example OWNER user;
```

退出PostgreSQL控制台

```
\q
```

查看数据库是否创建成功

```
psql -U user -d smoke_example -h 127.0.0.1 -p 5432
```

切回系统用户

```
su user
```

切换到smoke文件夹下

```
cd ...
```

建立数据库配置文件

```
echo "postgresql://user@localhost:5432/smoke_example" >db.url 或  
echo "postgresql://user:password@localhost:5432/smoke_example" >db.url
```

编译

```
deepdive compile
```

运行

```
deepdive run
```

运行过程中按提示输入密码，进入vi时退出即可

```
:wq
```

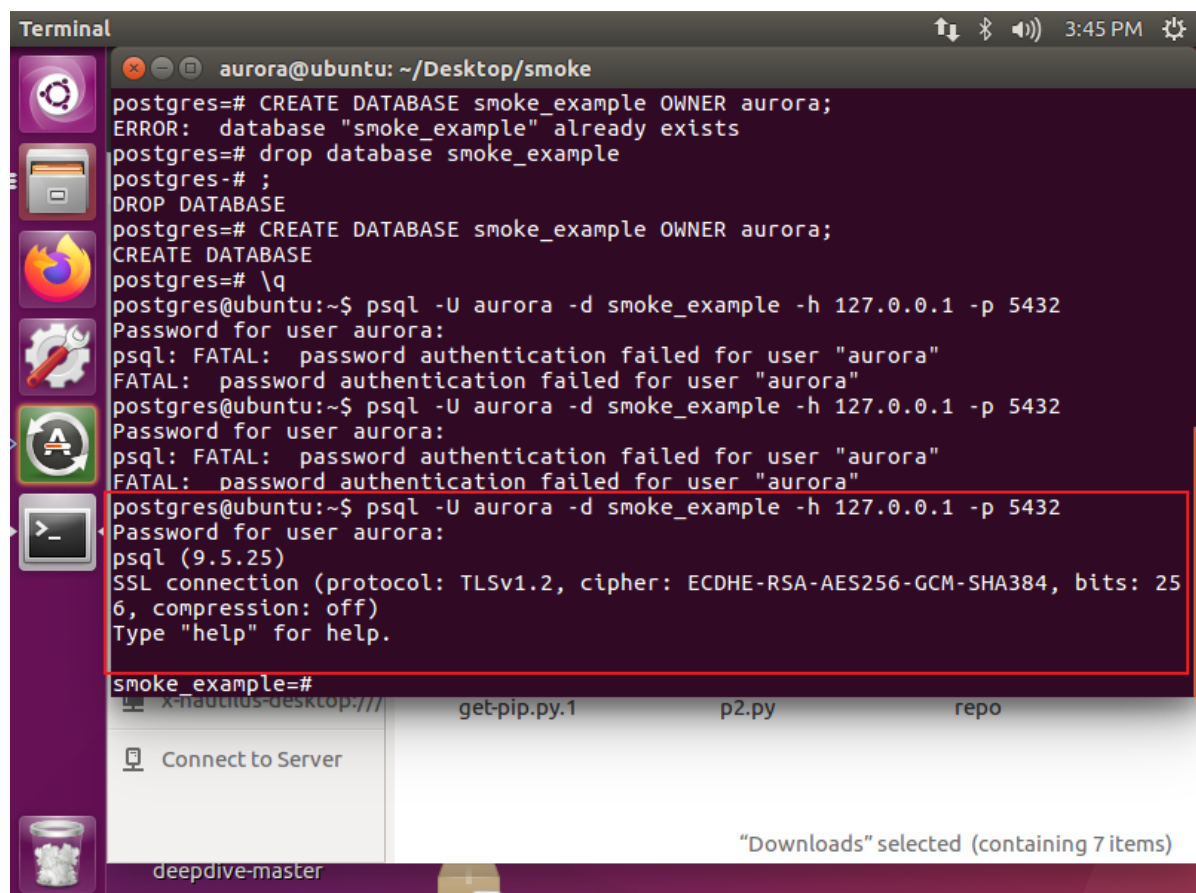
运行完成后查看结果

```
psql smoke_example  
\dv  
select * from tablename;
```

其余的安装过程没有展示，安装过程中存在一个问题，使用curl时会存在一些域名解析的问题，将域名放到对应的文件中，手动进行解析，就可以了。

查询数据库是否存在

发现此时已经有这个数据库了



The screenshot shows a terminal window titled 'aurora@ubuntu: ~/Desktop/smoke'. The user is in the PostgreSQL shell. The following commands and outputs are shown:

```
postgres=# CREATE DATABASE smoke_example OWNER aurora;  
ERROR: database "smoke_example" already exists  
postgres=# drop database smoke_example  
postgres=# ;  
DROP DATABASE  
postgres=# CREATE DATABASE smoke_example OWNER aurora;  
CREATE DATABASE  
postgres=# \q  
postgres@ubuntu:~$ psql -U aurora -d smoke_example -h 127.0.0.1 -p 5432  
Password for user aurora:  
psql: FATAL: password authentication failed for user "aurora"  
FATAL: password authentication failed for user "aurora"  
postgres@ubuntu:~$ psql -U aurora -d smoke_example -h 127.0.0.1 -p 5432  
Password for user aurora:  
psql: FATAL: password authentication failed for user "aurora"  
FATAL: password authentication failed for user "aurora"  
postgres@ubuntu:~$ psql -U aurora -d smoke_example -h 127.0.0.1 -p 5432  
Password for user aurora:  
psql (9.5.25)  
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 256, compression: off)  
Type "help" for help.  
smoke_example=#
```

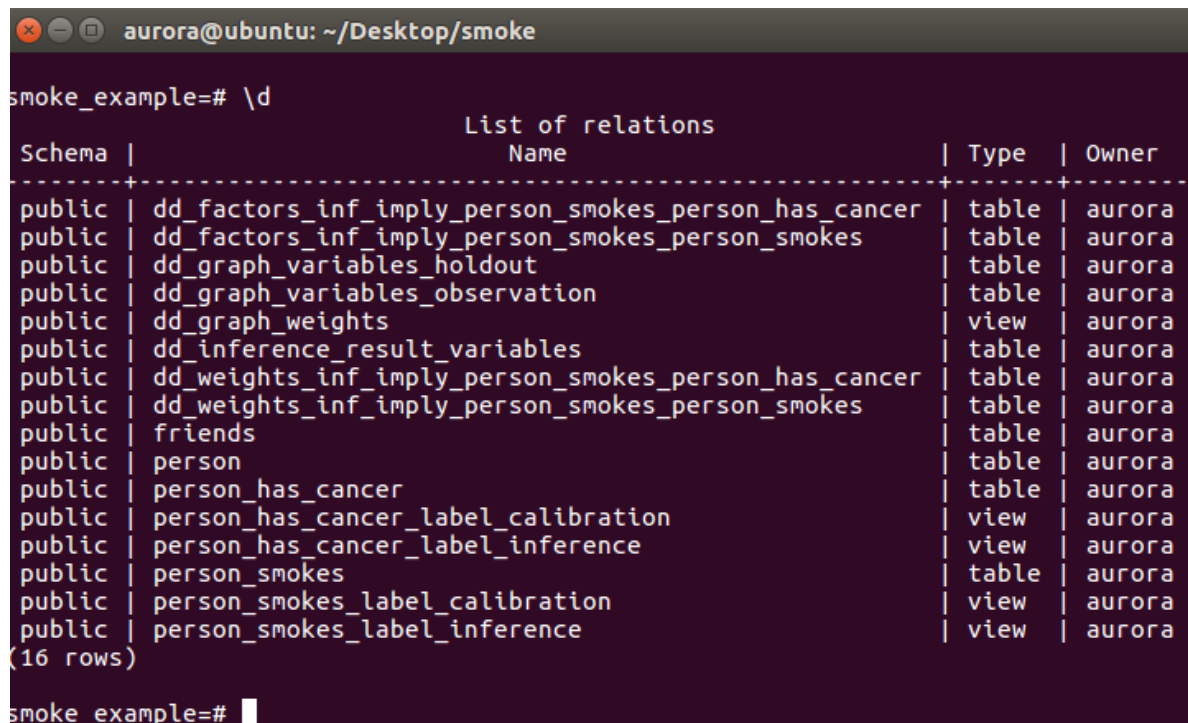
The last command prompt 'smoke_example=#' is highlighted with a red box, indicating the database exists.

使用deepdive

编译
deepdive compile

运行
deepdive run

使用deepdive进行操作后，发现数据库中存在了这些信息



```
aurora@ubuntu: ~/Desktop/smoke
smoke_example=# \d
```

Schema	Name	Type	Owner
public	dd_factors_inf_imply_person_smokes_person_has_cancer	table	aurora
public	dd_factors_inf_imply_person_smokes_person_smokes	table	aurora
public	dd_graph_variables_holdout	table	aurora
public	dd_graph_variables_observation	table	aurora
public	dd_graph_weights	view	aurora
public	dd_inference_result_variables	table	aurora
public	dd_weights_inf_imply_person_smokes_person_has_cancer	table	aurora
public	dd_weights_inf_imply_person_smokes_person_smokes	table	aurora
public	friends	table	aurora
public	person	table	aurora
public	person_has_cancer	table	aurora
public	person_has_cancer_label_calibration	view	aurora
public	person_has_cancer_label_inference	view	aurora
public	person_smokes	table	aurora
public	person_smokes_label_calibration	view	aurora
public	person_smokes_label_inference	view	aurora

(16 rows)

```
smoke_example=#
```

上图展示的内容即是本次实验使用deepdive进行知识抽取获得的结果，从原本的4个表变为了16个表。

四、实践结果

对其中的部分schema进行查看

person_smokes_label_inference

```
smoke_example=# select * from person_smokes_label_inference;
```

person_id	id	label	category	expectation
4	9		1	0.642
2	7		1	0.542
6	11		1	0.467
5	10		1	0.451

(4 rows)

dd_factors_inf_imply_person_smokes_person_has_cancer

```
smoke_example=# select * from
dd_factors_inf_imply_person_smokes_person_has_cancer;
person_smokes.R0.id | person_has_cancer.R1.id
```

-----+-----	
6	0
7	1
8	2
9	3
10	4
11	5

(6 rows)

dd_factors_inf_imply_person_smokes_person_smokes

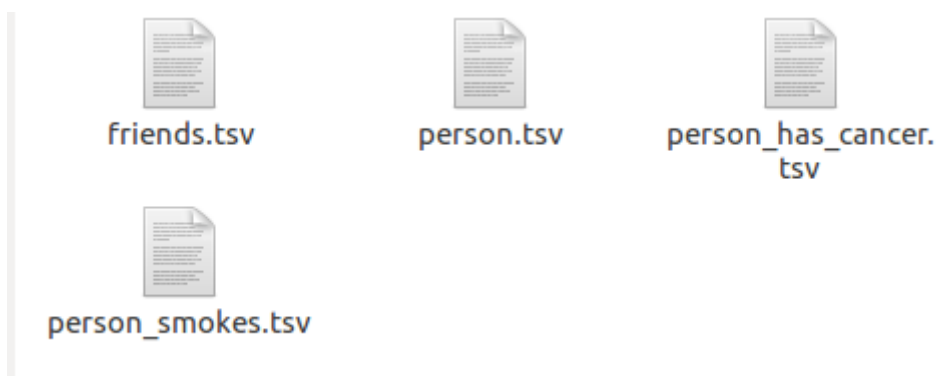
```
smoke_example=# select * from dd_factors_inf_imply_person_smokes_person_smokes;
person_smokes.R0.id | person_smokes.R1.id
```

-----+-----	
6	9
6	7
6	8
7	6
8	9
8	6
9	6
9	8
10	11
11	10

(10 rows)

五、实践心得

本次实验使用了deepdive工具对smoke数据进行了处理，smoke数据包含以下几个部分



经过deepdive对于数据以及其中关系的挖掘，可以获得更多的信息，甚至可以得到对于吸烟导致病症的概率。