# 实践二　　DeepDive关系抽取

课程名称：　知识工程　　实验日期：　2023/6/22

班级：　人工智能 x 班　姓名：　　学号：

## 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）

Ubuntu16

安装 python

```
root@ubuntu:/mnt/share# python
Python 2.7.12 (default, Jul 21 2020, 15:19:50)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more infor
>>>
KeyboardInterrupt
>>> quit()
root@ubuntu:/mnt/share# java
```

安装 java

```
root@ubuntu:/mnt/share# java -version
openjdk version "1.8.0_292"
OpenJDK Runtime Environment (build 1.8.0_292-8u292-b10-0ubuntu1~16.04.1-b10)
OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)
```

安装 CNdeepdive

```
Processing triggers for libc-bin (2.23-0ubuntu11.2) ...
+ 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
a
 1) deepdive
 2) deepdive_docker_sandbox
esac

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

# don't put duplicate lines or lines start
ive$ gedit ~/.bashrc
ive$ source  ~/.bashrc
ive$
```

安装 postgresql

```
+ 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.v884VYv
## Finished installation for postgres
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$
```

安装 nlp 环境

```
## Finished installation for postgres
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$ ./nlp_setup.sh
Install Dependency.
Denpendency Already Installed.
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$
```

项目环境搭建

实验一：

psql postgres

CREATE DATABASE smoke_example OWNER zq;

```
LINE 1: CREATE DATABASE smoke example OWNER CREATE DATABASE
                                              ^
postgres=# CREATE DATABASE smoke_example OWNER zq;
CREATE DATABASE
postgres=#
```

psql -U zq -d smoke_example -h 127.0.0.1 -p 5432

```
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$ psql -U zq -d smoke_example -h 127.0
.0.1 -p 5432
psql (9.5.25)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 25
6, compression: off)
Type "help" for help.

smoke_example=#
```

echo "postgresql://zq:zq@localhost:5432/smoke_example" > db.url

```
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$ echo "postgresql://zq@localhost:5432
/smoke_example">db.url
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive$
```

实验二：

psql smoke_example

```
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive/smoke$ psql smoke_example
psql (9.5.25)
Type "help" for help.

smoke_example=# \q
zq@ubuntu:/mnt/share/CNdeepdive/CNdeepdive/smoke$
```

psql transaction

CREATE DATABASE transaction OWNER zq;

psql -U zq -d transaction -h 127.0.0.1 -p 5432

echo "postgresql://zq@localhost:5432/transaction ">db.url

ALTER ROLE zq CREATEROLE SUPERUSER;

使用 deepdive 进行操作后，发现数据库中存在了这些信息

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

## 四、实践结果

### 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

实验使用 **deepdive** 工具来处理 **smoke** 数据。在 **deepdive** 对数据和关系进行挖掘后，可以获得更多的信息，甚至是关于吸烟导致疾病的概率，体会到了知识工程的艰难。