

PART 4

自然语言查询接口 的编程实现

Made by 倪浩垠

NL2SQL——大体思路

输入的自然语言：张三的家庭住址

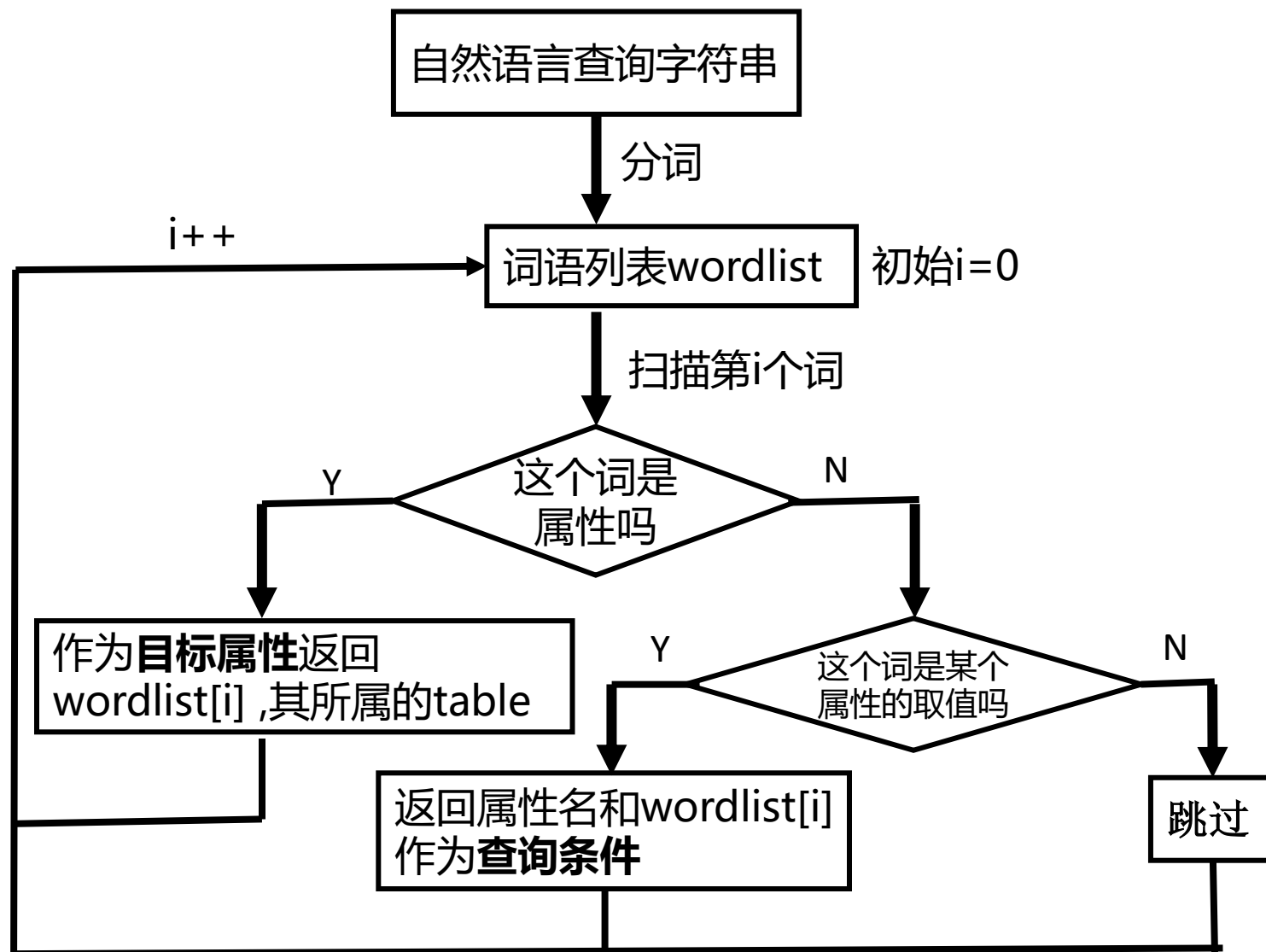
输出的SQL: select 家庭住址 from 职员信息表 where 姓名 = '张三'

目标：完成一道填空题，找出select的**目标属性**，from后的**table**和where后的**查询条件**

思路：扫描输入的自然语言，判断每个词属于哪种标签，将它填入相应的空格

参考论文：数据库汉语自然语言查询技术研究，王代星

NL2SQL——算法流程



张三的家庭地址

分词

张三/的/家庭住址



家庭住址：职员信息表里的属性

的：什么都不是

张三：是职员信息表里'姓名'属性的取值

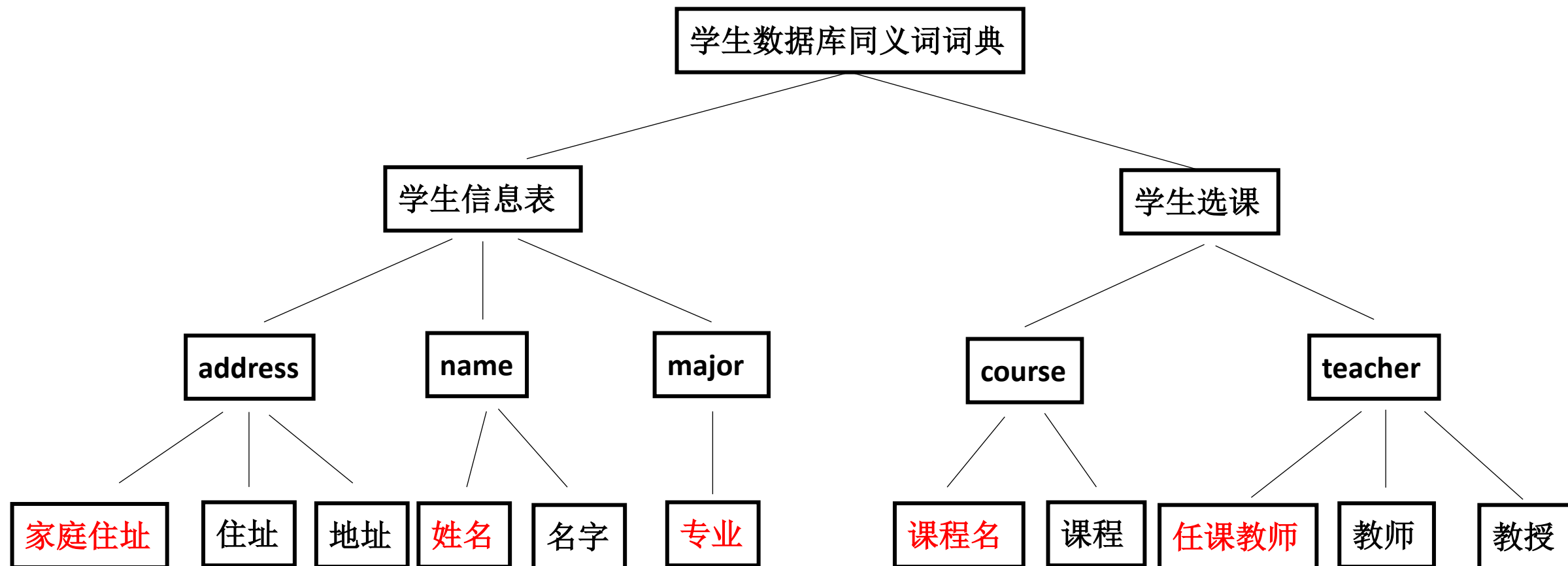
select 家庭住址

from 职员信息表

where 姓名 = '张三'

NL2SQL——同义词查询功能

使用同义词词典来实现同义词查询功能。假如扫描到的词为“教授”，那么搜索“教授”节点，将“教授”替换为“教授”的兄弟节点“任课教师”。



NL2SQL——总结

使用的python包：anytree（实现同义词词典），pymysql（与mysql对接），**jieba(中文分词)**，re(正则表达式)，tkinter(GUI)

缺点：

- 1 无法实现复杂的自然语言查询
- 2 查询的准确率极度依赖分词准确率

优点：

- 1 编程实现简单
- 2 部署方便
- 3 能实现简单的自然语言查询

展望：在NL2SQL这方面，我认为深度学习才是王道。





谢谢观看!