

《英雄联盟》职业和难度知识图谱

余凯(22121132), 鲁国山(22121120), 李立尧(22121103), 石昊楠(22121086)

1 项目介绍

《英雄联盟》（League of Legends，简称 LOL）是由美国拳头游戏（Riot Games）开发、中国内地由腾讯游戏代理运营的英雄对战 MOBA 竞技网游，是一款风靡全球的电子游戏。游戏里拥有数百个个性英雄，并拥有排位系统、符文系统等特色养成系统。作为杭州 2022 年第 19 亚洲运动会的电子竞技项目之一，《英雄联盟》亟需得到更多人的了解。

《英雄联盟》游戏中一共包含 157 个英雄，玩家在选择一个英雄后会与队友一起攻击另一支队伍。不同英雄的攻击能力、防御能力、玩家难度和购买需要的金币各不相同，且有划分不同的阵营和角色，甚至在丰富的背景故事中还有英雄间的亲属关系等。因此，本项目聚焦于构建英雄和其角色的实体关系知识图谱。其中实体包括英雄、角色；关系包括英雄与角色的关系（属于某一阵营）、英雄与英雄的关系（兄弟姐妹等）。

首先我们通过爬虫采集了英雄的基本数据，并对其预处理以储存成结构化数据。然后采用自定义的规则抽取和构建英雄实体间的关系，以构建知识图谱的三元组。我们通过 Python 将构建出的数据导入 Neo4j 的图数据库中，真正完成图谱的构建、存储和可视化。此外我们还构建了 Database-free 的可视化界面，方便直接预览数据。在基于 Neo4j 的图数据库上，我们实现了一个简单的智能问答机器人。

2 项目运行方式

- 配置要求：要求配置 Neo4j 数据库及相应的 Python 依赖包。Neo4j 数据库用户名密码记住，并修改相应文件。
- 知识图谱数据导入：python build_herosgraph.py。
- 启动问答：python chatbot_graph.py

3 项目效果

3.1 原始数据集

```
1  {
2      "hero": [
3          {
4              "heroId": "1",
5              "name": "黑暗之女",
6              "alias": "Annie",
7              "title": "安妮",
8              "roles": [
9                  "mage"
10             ],
11             "isWeekFree": "0",
12             "attack": "2",
13             "defense": "3",
14             "magic": "10",
15             "difficulty": "6",
16             "selectAudio": "https://game.gting.cn/images/lo/act/img/vo/choose/1.ogg",
17             "banAudio": "https://game.gting.cn/images/lo/act/img/vo/ban/1.ogg",
18             "isARAMweekfree": "0",
19             "ispermanentweekfree": "0",
20             "changeLabel": "无改动",
21             "goldPrice": "4800",
22             "couponPrice": "2000",
23             "camp": "",
24             "campId": "",
25             "keywords": "安妮,黑暗之女,火女,Annie,anni,heianzhinv,huonv,an,hazn,hn"
26         },
27         {
28             "heroId": "2",
29             "name": "狂战士",
30             "alias": "Olaf",
31             "title": "奥拉夫",
32             "roles": [
33                 "fighter",
34                 "tank"
35             ],

```

图 3-1 原始数据集

3.2 自定义可视化界面

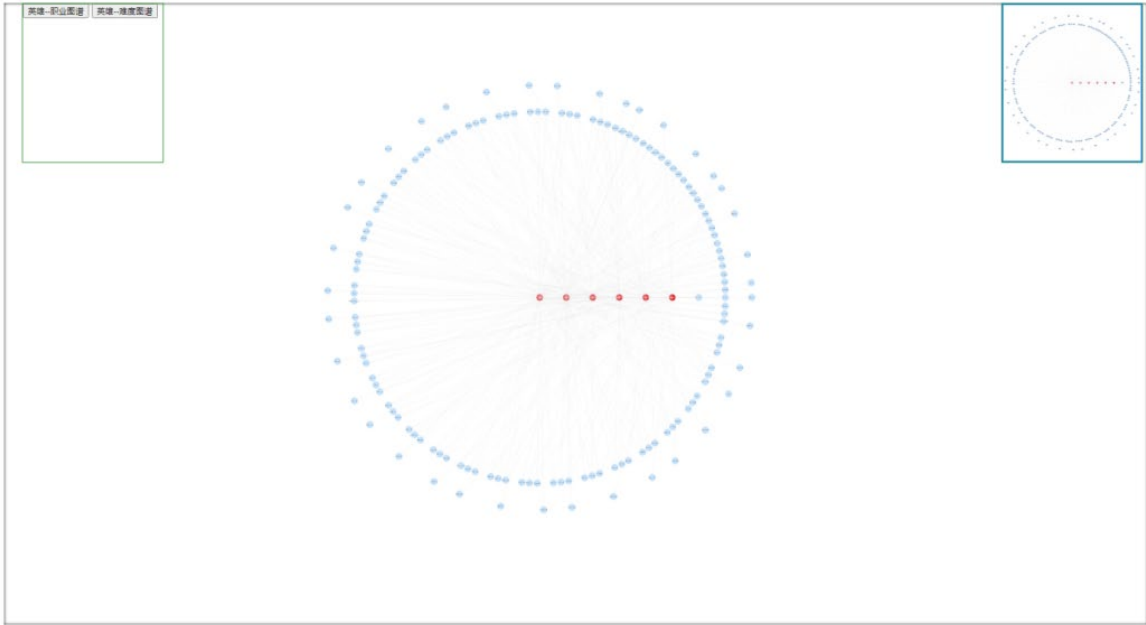


图 3-2 自定义可视化界面

3.3 Neo4j 可视化界面

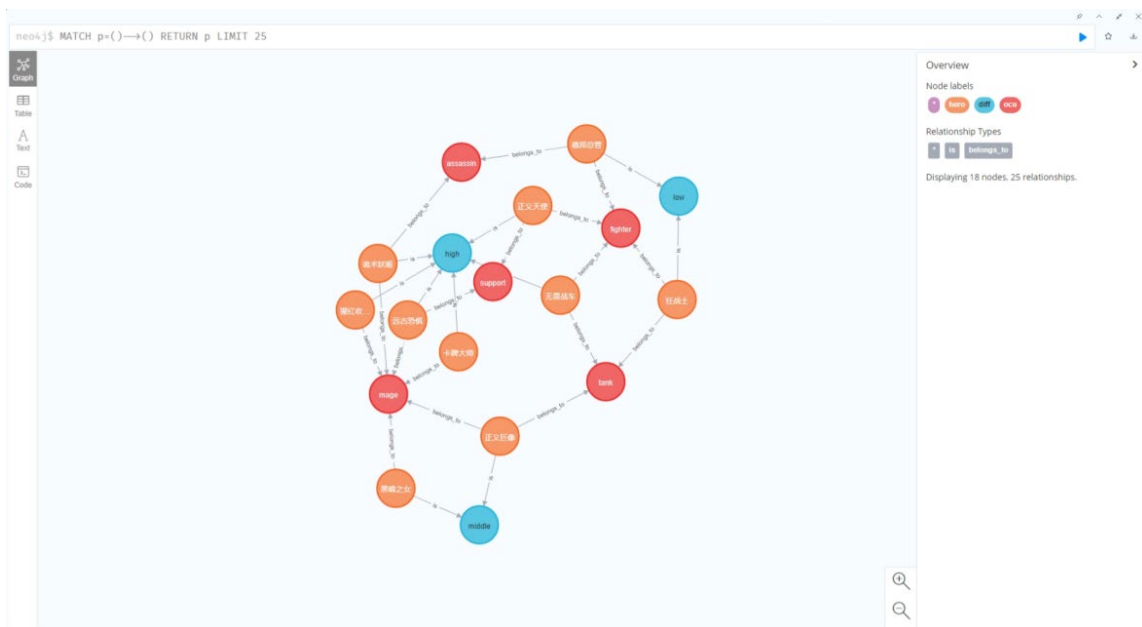


图 3-3 英雄--难度知识图谱

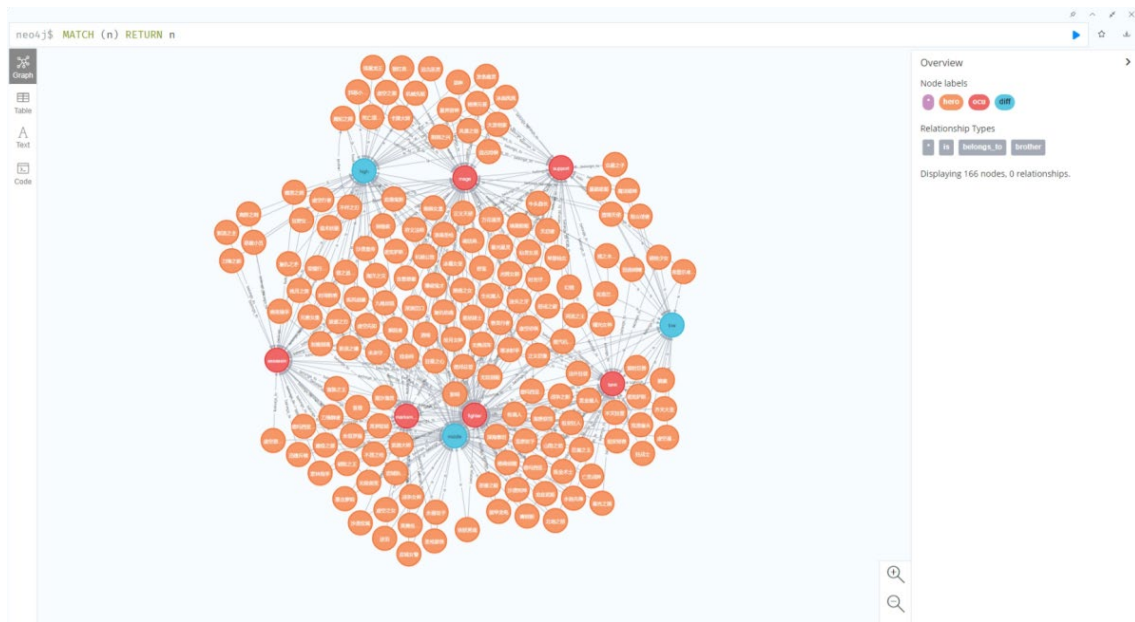


图 3-4 英雄--职业知识图谱

3.4 问答界面

```

Run: chatbot_graph x build_txt x
D:\ProgramData\Anaconda3\python.exe C:/Users/admin/Desktop/Heros/chatbot_graph.py
model init finished .....
用户: 刀锋之影难度大吗?
小TOT: 英雄“刀锋之影”的难度系数是7, 是high难度。
用户: 光辉女郎的兄弟姐妹是谁?
小TOT: 英雄“光辉女郎”的攻击点数是2, 防御点数是4, 难度是5, 需要花费3150个金币购买。
英雄“光辉女郎”的兄弟姐妹是英雄“德玛西亚之力”。
用户: 光辉女郎有兄弟姐妹吗?
小TOT: 有的。英雄“光辉女郎”的兄弟姐妹是英雄“德玛西亚之力”。
用户: mage都有谁?
小TOT: 角色“mage”下的英雄有: 远古恐惧, 仙灵女巫, 虚空先知, 皎月女神, 堕落天使, 暮光星灵, 远古巫灵, 愁云使者, 众星之子, 酒桶,
用户: mage都有谁?
小TOT: 角色“mage”下的英雄有: 远古恐惧, 仙灵女巫, 虚空先知, 皎月女神, 堕落天使, 暮光星灵, 远古巫灵, 愁云使者, 众星之子, 酒桶,
用户: mage下有几个人?
小TOT: 角色“mage”下共有61个英雄。
用户: |
  
```

图 3-5 问答界面

4 实现细节

4.1 爬取数据

数据使用 python 编写脚本，并从 <https://game.gting.cn> 英雄联盟官网 API 爬取，并将其保存为 JSON 数据格式文件，以方便后续的调用。

爬取脚本首先设定需要爬取的网页，然后填充 headers 头部变量以指示服务器使用什么浏览器进行登录爬取。然后调用 python 的 request 包，将整个英雄信息的 js 文件爬取下来，并使用 json 函数将其转化为 JSON 数据文件。

```
base_url='https://game.gting.cn/images/lol/act/img/js/heroList/hero_list.js'  
headers={'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.9 Safari/537.36'}  
  
response=requests.get(base_url,headers=headers)  
base_data=response.json()  
  
basedata=str(base_data)  
data_id=re.findall(r"heroId": '(.*?)', 'name",basedata)  
data_name=re.findall(r"name": '(.*?)', 'alias",basedata)  
data_alias=re.findall(r"alias": '(.*?)', 'title",basedata)  
data_title=re.findall(r"title": '(.*?)', 'roles",basedata)  
data_roles=re.findall(r"roles": '(.*?)', 'isWeekFree",basedata)  
data_attack=re.findall(r"attack": '(.*?)', 'defense",basedata)  
data_defense=re.findall(r"defense": '(.*?)', 'magic",basedata)  
data_magic=re.findall(r"magic": '(.*?)', 'difficulty",basedata)  
data_difficulty=re.findall(r"difficulty": '(.*?)', 'selectAudio",basedata)  
data_gold=re.findall(r"goldPrice": '(.*?)', 'couponPrice",basedata)  
data_coupon=re.findall(r"couponPrice": '(.*?)', 'camp",basedata)
```

转化出来的 JSON 文件如下图所示



```
1 {  
2   "hero": [  
3     {  
4       "heroId": "1",  
5       "name": "黑暗之女",  
6       "alias": "Annie",  
7       "title": "安妮",  
8       "roles": [  
9         "mage"  
10      ],  
11      "isWeekFree": "0",  
12      "attack": "2",  
13      "defense": "3",  
14      "magic": "10",  
15      "difficulty": "6",  
16      "selectAudio": "https://game.gting.cn/images/lol/act/img/vo/choose/1.ogg",  
17      "banAudio": "https://game.gting.cn/images/lol/act/img/vo/ban/1.ogg",  
18      "isARAMweekfree": "0",  
19      "ispermanentweekfree": "0",  
20      "changeLabel": "无改动",  
21      "goldPrice": "4800",  
22      "couponPrice": "2000",  
23      "camp": "",  
24      "campId": "",  
25      "keywords": "安妮,黑暗之女,火女,Annie,anni,heianzhinv,huonv,an_hazn,hn"  
26     },  
27     {  
28       "heroId": "2",  
29       "name": "狂战士",  
30       "alias": "Olaf",  
31       "title": "奥拉夫",  
32       "roles": [  
33         "fighter",  
34         "tank"  
35      ],  
36     }  
37   ]  
38 }
```

图 4-1 爬取数据

4.2 自定义可视化界面

4.2.1 环境配置

自定义的可视化界面主要是一个 demo 展示的功能，基于 Vue.js 前端框架和 G6 图可视化引擎而形成的 WEB 端的知识图谱可视化界面。

表 4-1 环境配置表

项目	技术
前端框架	Vue.js
可视化技术	G6 图可视化引擎
组件技术	Element-ui 技术

4.2.2 数据加载和展示

首先进行画布的初始化，将画布的模式设定为可拖动，节点可移动，默认环形布局，并设定画布中的默认节点形状、大小和连线参数等，具体功能设定如下图所示。

```
// 画布初始化
const graph = new G6.Graph({
  container: 'canvasContainer',
  width: document.getElementById('canvasContainer').clientWidth,
  height: document.getElementById('canvasContainer').clientHeight,
  // 开启拖拽功能，可以拖拽节点
  renderer: 'svg',
  // 交互模式
  // drag-mode: 支持拖拽节点
  // drag-combo: 支持拖拽组合
  // collapse-expand-combo: 支持组合的缩小与放大
  // drag-canvas: 支持画布的拖拽
  // zoom-canvas: 支持画布的放大与缩小画布
  modes: {
    default: [
      'drag-node',
      'drag-combo',
      'collapse-expand-combo',
      'drag-canvas',
      'zoom-canvas',
      'click-select',
      'brush-select'
    ],
  },
  linkCenter: true,
  layout: {
    type: 'concentric',
    center: [document.getElementById('canvasContainer').clientWidth/2, document.getElementById('canvasContainer').clientHeight/2],
    linkDistance: 500, // 可选，默认
    preventOverlap: true, // 可选，必须配合 modeSize
    modeSize: 10, // 可选
    sweep: 20, // 可选
    equidistant: false, // 可选
    startAngle: 0, // 可选
    clockwise: false, // 可选
    minNodeSpacing: 5, // 可选
    maxLevelDiff: 1, // 可选
    sortBy: 'degree', // 可选
    workerEnabled: true // 可选，开启 web-worker
  },
});
```

图 4-2 数据加载代码

其次，进行数据的提取，生成英雄--职业和英雄--难度对应的数据图。其次，使用数据图进行节点的生成，并通过对数据的划分形成类型不同的节点，最后通过节点之间对应关系进行连线。其中，生成的英雄--职业和英雄--难度如下图所示。

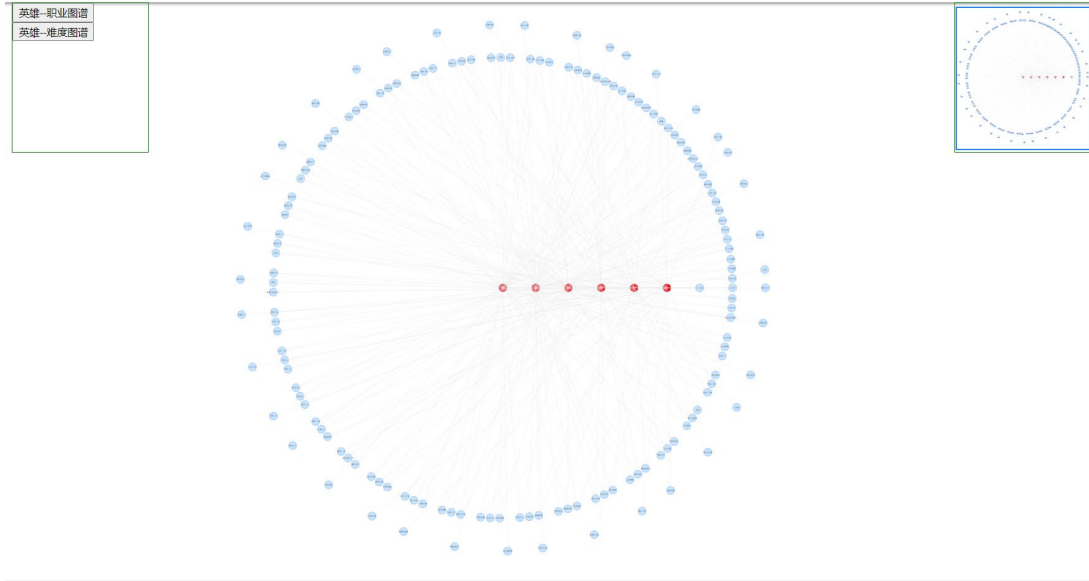


图 4-3a 英雄--职业知识图谱

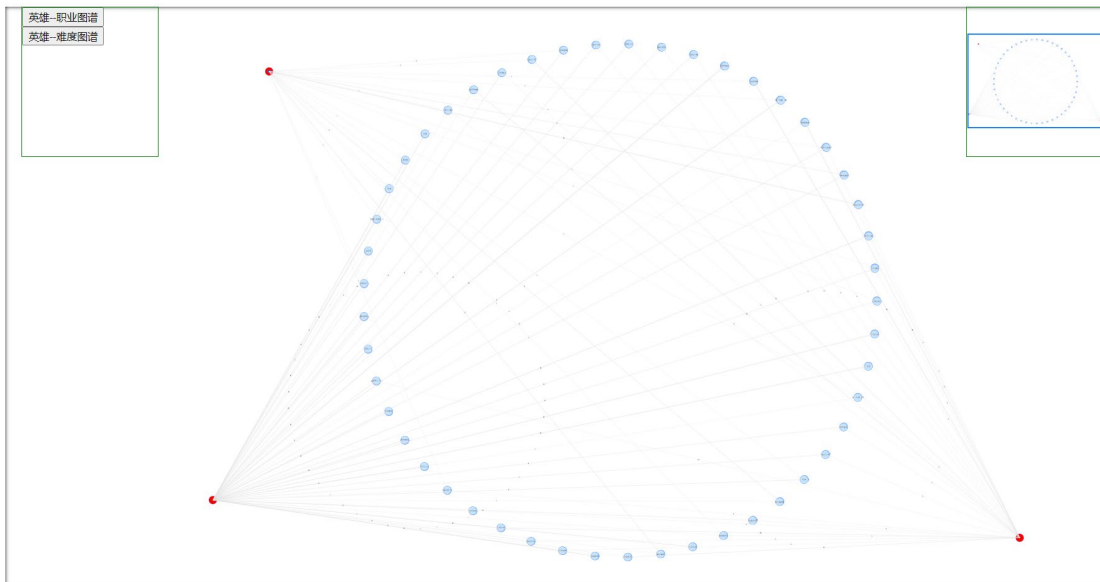


图 4-3b 英雄--难度图谱

4.3 知识图谱构建

使用 Neo4j, Neo4j 的可视化界面。全图效果展示：

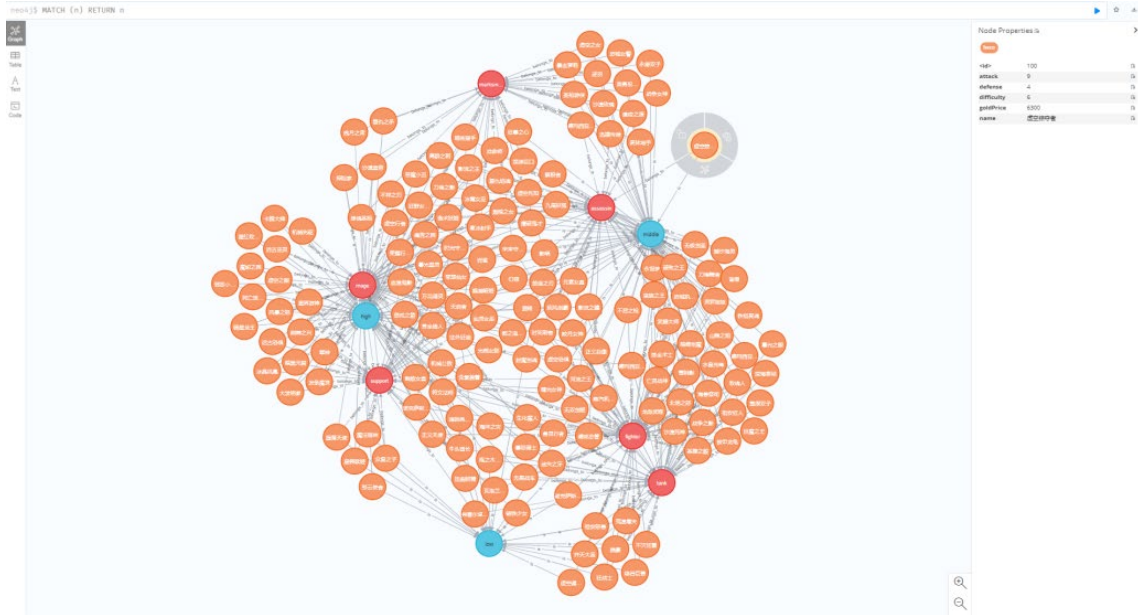


图 4-4 全图效果

其构建方法如下：

- (1) 首先安装 `java11` 以及 `Neo4j` 社区版的图数据库，并导入 `py2neo` 包。
- (2) 初始化与 `Neo4j` 数据库的连接。
- (3) 读取爬取的 `json` 数据，并一一解析出需要构建的实体内容和属性以及构建相应的关系。如英雄和职业、难度之间的实体关系，英雄与英雄之间的兄妹关系等等。以及每个实体的属性信息构建。
- (4) 用“`Cypher`”语言创建对应的 `Neo4j` 图数据库中的实体节点和关系边。

4.3.1 知识图谱实体类型

表 4-2 实体类型表

实体类型	中文含义	实体数量	举例
Hero	英雄	157	机械先驱
Occupation	职业	6	Mage
Difficulty	难度	3	High
Total	总计		

4.3.2 知识图谱实体关系类型

表 4-3 实体关系类型表

实体关系类型	中文含义	关系数量	举例
belongs_to	属于	277	机械先驱属于 Mage（法师）职业
is	难度是	157	机械先驱难度是 High（高操作难度）
brother_of	是兄弟姐妹	4	机械先驱无兄弟姐妹
Total	总计	438	

4.3.3 知识图谱属性类型

表 4-4 属性类型表

属性类型	中文含义	举例
name	英雄名称	机械先驱
attack	攻击点数	9
defense	防御点数	4
difficulty	玩家难度	6

4.4 自动问答

本项目的智能问答机器人基于预设定的问答模板识别进行回答。首先会根据用户提的问题去判断问句类型（如是介绍英雄还是检索英雄角色）；接着根据问句类型执行不同的数据库检索语句检索对应的数据；最后根据问句类型和检索结果合并到回答模板中并输出。

4.4.1 脚本结构

表 4-5 脚本结构表

文件	脚本
question_classifier.py	问句类型分类脚本
question_parser.py	问句解析脚本
chatbot_graph.py	问答程序脚本

4.4.2 支持问答类型

表 4-6 支持问答类型表

问句类型	中文含义	问句举例
hero_desc	英雄介绍	刀锋之影是谁？
hero_role	英雄角色	刀锋之影是什么角色？
hero_attack	英雄攻击点数	刀锋之影攻击强吗？
hero_defense	英雄防御点数	刀锋之影防御强吗？
hero_difficulty	英雄难度点数	刀锋之影难度大吗？
hero_brother	英雄兄弟姐妹	光辉女郎的兄弟姐妹是谁？

问句类型	中文含义	问句举例
has_brother	英雄是否有兄弟姐妹	光辉女郎有兄弟姐妹吗？
role_cluster	角色聚类	mage 都有谁？
role_count	角色下英雄统计	mage 下有几个人？

4.4.4 问答效果展示

```

*****
model init finished .....
用户: 你好
小TOT: 您好,我是英雄联盟职业难度咨询助理--小TOT,希望可以帮到您。祝您游戏愉快!

用户: 狂战士是谁
小TOT: 英雄“狂战士”的攻击点数是9,防御点数是5,难度是3,需要花费1350个金币购买。

用户: 正义巨像是什么角色
小TOT: 英雄“正义巨像”的攻击点数是1,防御点数是10,难度是5,需要花费3150个金币购买。
英雄“正义巨像”的职业是tank。英雄“正义巨像”的职业是mage。

用户: 卡牌大师属于哪个职业
小TOT: 英雄“卡牌大师”的职业是mage。

用户: 德邦总管攻击强吗
小TOT: 英雄“德邦总管”的攻击点数是8。

用户: 德邦总管防御多少
小TOT: 英雄“德邦总管”的防御点数是6。

用户: 刀锋之影难度大吗?
小TOT: 英雄“刀锋之影”的难度系数是7,是high难度。

用户: 光辉女郎的兄弟姐妹是谁?
小TOT: 英雄“光辉女郎”的攻击点数是2,防御点数是4,难度是5,需要花费3150个金币购买。
英雄“光辉女郎”的兄弟姐妹是英雄“德玛西亚之力”。

用户: 光辉女郎有兄弟姐妹吗?
小TOT: 有的。英雄“光辉女郎”的兄弟姐妹是英雄“德玛西亚之力”。

用户: mage都有谁?
小TOT: 角色“mage”下的英雄有:远古恐惧,仙女女巫,虚空先知,皎月女神,堕落天使,曙光星灵,远古巫灵,愁云使者,众星之子,酒桶,邪恶小法师,发条魔灵,狂野女猎手,风暴之怒,黑暗之女,天启者,正义巨像,符文法师,诺克萨斯统领,虚空行者,翠神,虚空恐惧,魔法猫咪,猩红收割者,解脱者,星籁歌姬,机械公敌,含羞蓓蕾,死亡颂唱者,诡术妖姬,黯黑元首,虚空之眼,冰晶凤凰,荆棘之兴,殇之木乃伊,九尾妖狐,时光守护者,戏命师,狂暴之心,惩戒之箭。

用户: mage下有几个人?
小TOT: 角色“mage”下共有61个英雄。
*****

```

图 4-5 问答效果展示

参考资料

本项目参考了项目《[医疗知识图谱的构建与自动问答](#)》。