

英雄联盟宇宙知识图谱

葛智超（22121036），魏浩文（22121177）

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。经过多年的发展，围绕英雄联盟这一IP已近发展出很多周边产业，英雄联盟宇宙便是其中之一。英雄联盟宇宙汇集了英雄们在一片名叫符文之地的魔法大陆上的传奇故事，大到世界观，小到地方的风土人情。我们使用Python爬取英雄们的背景故事，对获取的数据进行数据清洗、抽取等处理，再把得到的三元组读入Neo4j，并搭建了简单的查询服务。

2 数据获取预处理

2.1 通过爬虫获取原始数据

经过调研，我们选择英雄联盟宇宙官方平台（https://yz.lol.qq.com/zh_CN/）作为主要数据来源。通过发送HTTP请求获取网站的JSON数据，然后对数据进行解析，得到原始数据。获取JSON的方法如下面的代码所示，我们主要获取英雄和他们所在地区的数据：

```
1  def get_json(url):
2      headers = {'user-agent': 'Mozilla/5.0'}
3      response = requests.get(url, headers=headers)
4      response.encoding = 'utf8'
5      d_json = response.json()
6      return d_json
7  def get_champions_info(city_url):
8      .....
9  def get_city_info(city_url):
10     .....
11
```

2.2 数据处理结果

原始数据中，结构化的数据如英雄的称号、定位等可以直接构建三元组导入知识图谱，而他们之间的关系则需要从背景故事文本抽取，得到初步的关系图谱，然后人工进行校正补全。我们针对英雄联盟宇宙给出了朋友、敌对、师徒、同一阵营等十余种关系。

除了以英雄为主体构建的三元组，我们还基于英雄联盟宇宙中的各个国家和地区建立了三元组，分析了各个城邦的风土人情，让知识图谱描绘的英雄联盟宇宙更接近官方描述。我们将获取的三元组存入csv文件，便于Neo4j读取。

经过数据处理，我们共得到四组结构化数据，分别是英雄实体及其属性、英雄关系、城邦数据和各城邦的风土人情数据。

57	凯南	狂暴之心	艾欧尼亚	法师	约德尔人	2010/4/8	历史悠久的均衡
58	卡兹克	虚空掠夺者	虚空之地	刺客	虚空生物	2012/9/27	虚空在成长，虚
59	千珏	永猎双子	符文之地	射手	神	2015/10/14	千珏，作为象征
60	克烈	暴怒骑士	诺克萨斯	战士	约德尔人	2016/8/10	无畏而且狂暴的
61	克格莫	深渊巨口	虚空之地	射手	虚空生物	2010/6/24	从艾卡西亚废墟
62	乐芙兰	诡术妖姬	诺克萨斯	刺客	人类	2010/11/2	即使是在秘密团
63	李青	盲僧	艾欧尼亚	战士	人类	2011/4/1	李青是艾欧尼亚
64	蕾欧娜	曙光女神	巨神峰	坦克	人类	2011/7/13	蕾欧娜是被灌注
65	莉莉娅	含羞蓓蕾	艾欧尼亚	战士	非人	2020/7/7	羞怯难当的仙灵
66	丽桑卓	冰霜女巫	弗雷尔卓德	法师	寒冰血脉	2013/4/30	大多数人都认为
67	卢锡安	圣枪游侠	符文之地	射手	人类	2013/8/22	卢锡安，一名光
68	璐璐	仙灵女巫	班德尔城	辅助	约德尔人	2012/3/20	璐璐是一位约德
69	拉克丝	光辉女郎	德玛西亚	法师	人类	2010/10/19	拉克珊娜·冥卫出
70	墨菲特	熔岩巨兽	以绪塔尔	坦克	非人	2009/9/2	墨菲特是一个庞
71	玛尔扎哈	虚空先知	虚空之地	法师	虚空生物	2010/6/1	玛尔扎哈是一名
72	茂凯	扭曲树精	暗影岛	坦克	非人	2011/2/16	茂凯是一只内心
73	易	无极剑圣	艾欧尼亚	刺客	人类	2009/2/21	易大师锤炼身体
74	厄运小姐	赏金猎人	比尔吉沃特	射手	人类	2010/9/8	以美貌闻名，但
75	孙悟空	齐天大圣	艾欧尼亚	战士	瓦斯塔亚	2011/7/26	悟空是一个瓦斯
76	莫德凯撒	铁铠冥魂	诺克萨斯	战士	非人	2010/2/24	两度被杀，三度

图1 英雄实体及属性

17	艾尼维亚	领导	艾希
18	艾尼维亚	朋友	努努和威朗普
19	厄斐琉斯	同一阵营	黛安娜
20	厄斐琉斯	朋友	蕾欧娜
21	艾希	妹妹	瑟庄妮
22	艾希	恋人	泰达米尔
23	艾希	领导	布隆
24	艾希	妹妹	丽桑卓
25	奥瑞利安·索尔	敌对	索拉卡
26	奥瑞利安·索尔	敌对	蕾欧娜
27	奥瑞利安·索尔	敌对	黛安娜
28	阿兹尔	领导	内瑟斯

图2 英雄关系三元组

	city_name ▼	city_slug ▼	city_context ▼
	比尔吉沃特	bilgewater	在远离大陆的蓝焰群岛边缘，坐落着独一无二的... 虽然比尔吉沃特是极其危险的地方，但这里也... 近来，上一任“强盗之王”遭到罢黜，这座城正如
	艾欧尼亚	ionia	在凶险海域环绕中，诸多同盟省份在一片被称... 虽然这片土地上的魔力可能变幻无常，栖息着... 但那一切都在十二年前结束了，因为那一年，... 如今的艾欧尼亚正处于脆弱的和平中。对于那... 艾欧尼亚的命运正是在微妙的平衡中，几乎无
	以绪塔尔	ixtal	以绪塔尔以精通元素魔法而闻名，也是第一批... 但是以绪塔尔的法师们将荒野作为屏障隔离了... 现在，于丛林中隔绝世界千年之后，以绪奥肯
	虚空之地	void	伴着宇宙的诞生，一声尖叫宣示了虚空的开端... 凡人被这种力量触碰，就等同于窥到了永恒的
	诺克萨斯	noxus	诺克萨斯是一个威名震天的强大帝国。在诺克... 古代诺克西人是残暴的野蛮人部落联合，他们... 无论社会立场、身世背景、祖国故乡和个人财... 虽然诺克萨斯有贤能统治的政治理想，但老一
	巨神峰	mount-targon	巨神峰是符文之地的世界之巅，这座高耸入云
	德玛西亚	demacia	德玛西亚是一个法理至上的强大王国，战功赫... 然而现在，王国中已经出现了变数。 德玛西亚雄都以禁魔石——一种可以抑制魔法

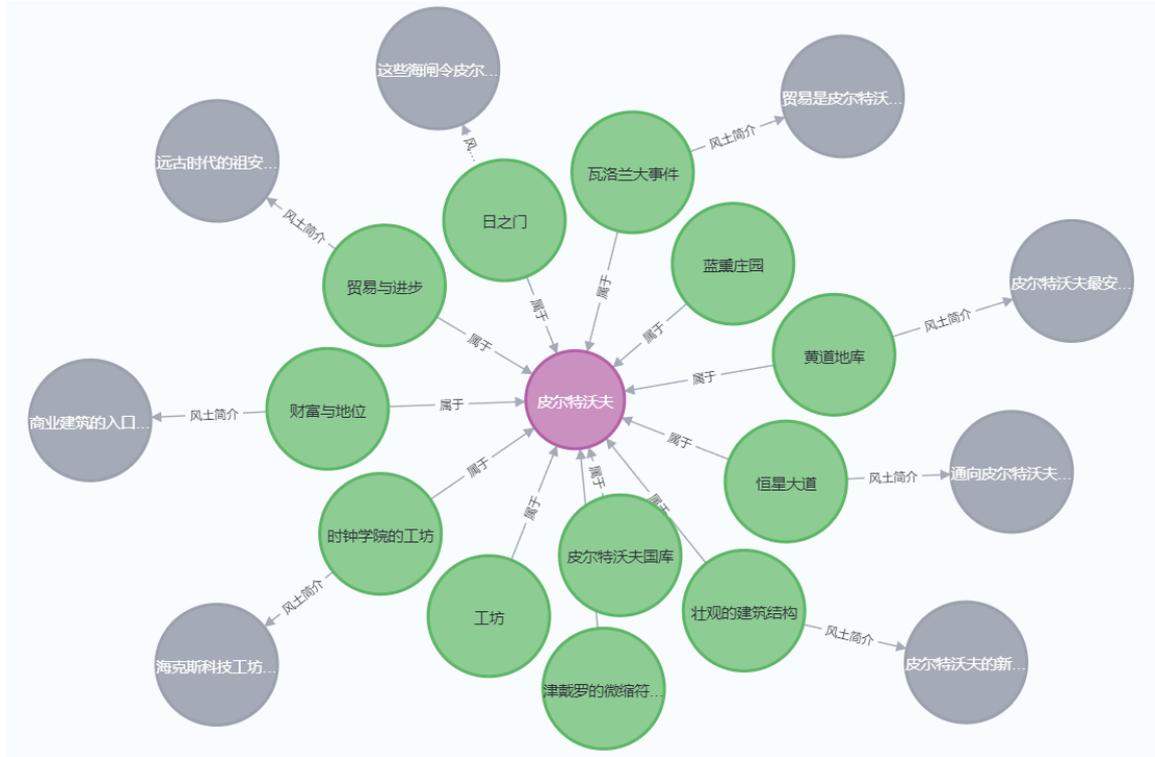
图3 城邦数据

蟒行群岛	瓦洛兰大陆上的人所说的蓝焰群岛，在水生水长的芭茹人的概念中	比尔吉沃特
娜伽卡波洛丝	芭茹文化的中心是娜伽卡波洛丝——代表生命、生长和永恒运动的	比尔吉沃特
密银城	一座坐落于德玛西亚西北岩石高地的城市	德玛西亚
黎明城堡	嘉文三世的宫殿	德玛西亚
英勇之厅	阵亡的德玛西亚将士们在此被人铭记	德玛西亚
光明使者神殿		德玛西亚
德玛西亚城的宏	统治者在人们雷鸣般的称颂声中接受任命	德玛西亚
科技魔法学院的	升向皮尔特沃夫	祖安
后街的海克斯科		祖安
呼气站	通往下方微光酒吧的入口	祖安
边境集市兴旺于		祖安
祖安城中处处生		祖安
津戴罗的微缩符		皮尔特沃夫
黄道地库	皮尔特沃夫最安全的银行	皮尔特沃夫
皮尔特沃夫国库		皮尔特沃夫

图4 各城邦的风土人情数据

3.3 各城邦的风土人情

通过我们构建的知识图谱，我们还可以了解到英雄联盟宇宙各城邦的风土人情：



4 总结与展望

在本次课程项目中，我们学习了从获取原始数据、预处理、抽取到构建知识图谱的过程。限于抽取技术的运用及官方平台的数据完整性，知识图谱不是很完整，这也是我们下一步工作打算继续学习的方向。