开题报告

- 1. 打算做什么: 打算做一个知识图谱,是一个关于春秋战国时期历史的图谱。 其中包含了人物关系,人物信息和事件信息。其中人物关系即包含了同一时期人物的关系,比如包含了齐桓公与管仲的君臣关系,也包含了不同时期人物关系,例如秦国的秦庄襄王到秦始皇的连续几代的跨越关系。人物信息包含了在位时间,姓、氏和谥号等具体信息,便于历史学者有跨度的研究问题。事件信息包含了战争信息,会盟信息等等信息,并在大事件中包含了子事件,有嵌入的关系,构成了不同国家不同君王以及不同名臣之间最直接的关系。 打算从时间跨度和空间跨度上做一个围绕春秋战国时期的一个历史知识图谱。
- 2. 为什么做这个题目: 这个知识图谱跨越了空间和时间, 而且包含了具体细节,对于研究历史和对历史感兴趣的朋友来说,即使查询具体资料的数据库,又是能把各类事件各类人物联系在一起的图谱,会减少历史学者的工作量,而且能让对历史感兴趣的朋友更加系统全面的接触历史,甚至是从中发现新的模式。
- 3. 为什么已有的产品不能解决问题:已有的产品大多是讲历史数据录入后的总结与查询工作,但是关系需要依靠可视化和综合全面的看待,才能发掘新的模式与规律。最重要的是目前据了解还没有类似的项目利用图数据库进行知识图谱的编辑。
- 4. 打算用什么数据,如何获取这些数据:用的是历史的数据,包括前文所提及的人物关系,人物信息和事件信息三大重点类的数据,其中数据已经经过获取。具体展示部分片段如下图所示:

事件编码	战争名称	时间起	时间止			主攻方	主守方	主攻-帮手	主守-帮手	主攻方-兵力
W1	长勺之战	684		长勺	(4	齐国-齐桓公	鲁国-鲁庄公			
W2	干时之战	685		干时		鲁国-鲁庄公	齐国-齐桓公			
W3	乘丘之战	684		乘丘((今	宋国-宋闵公;齐国-3	鲁国-鲁庄公			
W4	泓水之战	638		泓水		楚国-楚成王	宋国-宋襄公	郑国	卫国、许国	、滕国
W5	韩原之战	645		韩原		秦国-秦穆公、丕豹、	晋国-晋惠公			
W6	城濮之战	632		城濮	(委	晋国-晋文公	楚国-令尹子玉	齐国、秦国、	陈国、蔡国	三万两千
W7	殽之战	628		殽山		晋国-晋襄公	秦国-孟明视			
W8	颖北之战	599		颖北		晋国-士会	楚国-楚庄王			
W9	邲之战	597		邲地		晋国-荀林父	楚国-楚庄王	士会、赵朔、	孙叔敖、子	反、子重
W10	鞍地之战	589		鞍地		齐国-齐顷公	晋国-郤克		鲁国、卫国	
W11	鄢陵之战	575		鄢陵		楚国	晋国	子重、子反、	栾书、郤至	
W12	三驾之战	572		郑国		晋国-晋悼公	楚国-楚共王			
W13	湛阪之战	557		湛阪		晋国-中行偃	楚国-公子格			
W14	长岸之战	525		长岸		吴国-公子光	楚国-子鱼、子	暇		
W15	柏举之战	506		柏举		吴国-阖闾	楚国-楚昭王	蔡国、唐国		3万
W16	五战及郢	511	506	楚国		吴国-阖闾	楚国-楚昭王	蔡国、唐国		
W17	夫椒之战	494		夫椒	j	越国-勾践	吴国-夫差			
W18	艾陵之战	484		艾陵		吴国-夫差	齐国-齐简公	鲁国、陈国、	邾国	10万
W19	笠泽之战	478		笠泽		越国-勾践	吴国-夫差			2万
W20	晋阳之战	454		晋阳		智伯	赵襄子	韩康子、魏恒	子	
W21	垂沙之役	301		垂沙		齐国-匡章;魏国-公孙	楚国-唐蔑			
		·盟事件 序列事件								

- 5. 数据量有多大: 三类加起来大概有 1000 行左右,每行有各种的语言信息,一 共有很多列的信息,信息十分纷繁复杂,十分冗杂,缺失值较多,但是缺失 值不是必要属性,只是某些细节刻画的缺失,在知识图谱中可以省略。
- 6. 数据可以直接使用,进行知识图谱的制作。
- 7. 数据具体处理,展示最难以处理的部分进行解释,如下图所示:

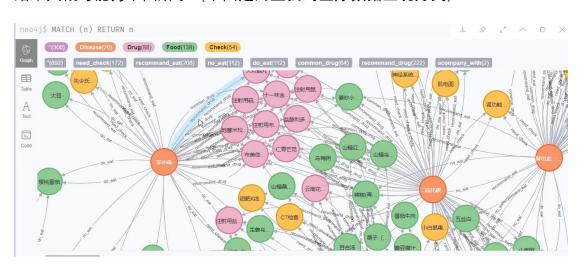
原因(借口)	意义	嵌套事件									
为了帮助太子郑,	正式成为春秋第一	个霸主									
为了平息宋国内乱	第一次以诸侯身份	主持多国会员	盟								
鲁国割地求和并入	盟	E5									
楚国要与齐国讲和	齐楚第一次交锋,	齐胜,楚答	应向周天:	子纳贡。							
宋襄公想称霸		E41									
城濮之战之后晋国	晋国经城濮之战,	一战而霸。									
晋楚常年争霸,导	结束了晋楚两个的	长期军事对峙	時,换来5	卢原60多年	的和平。						
邾、滕、薛、杞、/	晋国实力的顶端。										
齐国想拉拢日益强	孔子作为鲁国大司	寇,组织这5	汤会盟 , 护	羊与齐国君	臣斗智斗	勇,实现了	'有理、有	利、有节的	勺争斗。是	:孔子在仕;	金上的巅峰。
吴国保全周王氏	吴国得到老牌强国	的认可,可证	胃称霸中原	京。为了展	示吴国的	实力,将至	:国精兵都	带过去展为	下,也给越	国偷袭的机	几会。
	勾践称霸, 成就春	秋最后一位军	荫主。								

其中原因,借口包含的是长句信息,我需要考虑到对信息的提取,如果实现过程比较顺利的情况下,我可能会考虑用自然语言处理的办法提取数据,将关系刻画的更加细致,如果自然语言处理效果不佳,我可能会选择直接将具

体信息与事件绑定不做额外分析。其他的数据关系都是有规律可循的,可以直接提取。其中数据分析主要考虑数据的储存和分类,尽量简介明白的表示信息之间的关系。

8. 最后预期的结果与呈现方式:

结果大概可能为下图所示(下图是网上找到医疗数据呈现方式)



呈现方式: 打算用 python 语言实现 Neoj4 数据库的 web 可视化开发。

工具准备:

1.后端: neo4j 图数据库储存网络节点与节点间的关系。

Web 框架 flask,一个基于 python 的 web 微框架。

(时间顺利会考虑用 Django 实现)

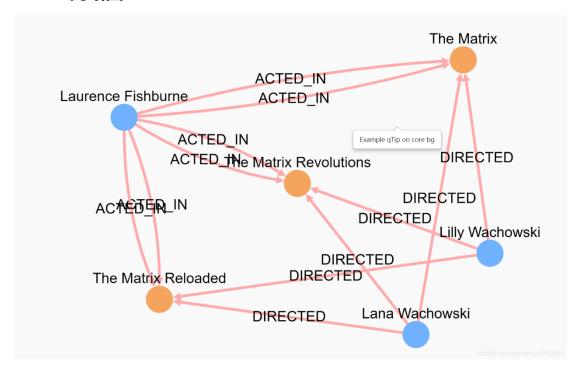
Py2neo 库和 The Neo4j Drivers for Python, python 操作 neo4j 的 API

2.前端: cytoscape.js 显示节点及节点间的关系

jQuery.js AJAX 必用的库

做出这种感觉

Movie网络图



对结果进行解释:各个元素之间的关系明了,可解释性强。