

OpenKG新冠知识图谱介绍

“2020年来，一场突如其来的新冠疫情席卷全球，在全民抗“疫”的关键时刻，OpenKG推出了新冠疫情专题知识图谱数据集，并在此基础上将数据转化成RDF数据，并基于gStore建立了知识图谱系统，为广大科研工作者提供知识服务”

本次在OpenKG所发布的新冠知识图谱数据基础上，将其转化为RDF三元组数据格式并导入gStore图数据库，提供基于云端的Endpoint接口服务；目前用户可以通过SPARQL查询语言进行访问；同时我们正在开发基于关键词等智能化访问接口，供大家学习和研究使用。

OpenKG开放的新冠知识图谱数据，总共包括8大类共17个知识图谱，分别为新冠科研图谱，新冠百科图谱，新冠健康图谱，新冠临床图谱，新冠英雄图谱，新冠热点事件图谱，新冠流行病学图谱和新冠物资图谱。

01 新冠科研图谱

科研图谱主要是从病毒分类、病毒基本信息、抗病毒药物、病毒亲缘关系图谱等角度进行科研分析，具体包括四个子图谱：

· 病毒分类图谱

病毒分类图谱是根据最新ncbi网站内容更新了病毒族系相关的科研基础数据，其中以新冠病毒科为主，主要是对病毒按照门、纲、目、科、属等进行分类分析。病毒分类图谱包括中文知识图谱和英文知识图谱，共两个知识图谱，名称为病毒分类图谱-中文、病毒分类图谱-英文。

三元组数量：1934954

实体数量：205633

谓语数量：16

· 新冠基本信息图谱

从新冠病毒的基因蛋白，宿主等相关信息出发构图。新冠基本信息图谱包括中文知识图谱和英文知识图谱，共两个知识图谱，名称为新冠基本信息图谱-中文、新冠基本信息图谱-英文。

三元组数量：2879

实体数量：553

谓词数量：10

· 抗病毒药物图谱

依据抗病毒药物、病毒、病毒相关蛋白和宿主及宿主蛋白间的关系构建图谱。抗病毒药物图谱包括中文知识图谱和英文知识图谱，共两个知识图谱，名称为病毒药物图谱-中文、病毒药物图谱-英文。

三元组数量：61796

实体数量：7999

谓语数量：2

· 病毒亲缘关系图谱

整理短时间内COVID-19病毒株的变异方向与地理位置分布数据包括中文知识图谱和英文知识图谱，共两个知识图谱，名称为病毒亲缘关系图谱-中文、病毒亲缘关系图谱-英文。

三元组数量：2237

实体数量：464

谓语数量：17（括号里面的为英文谓词）

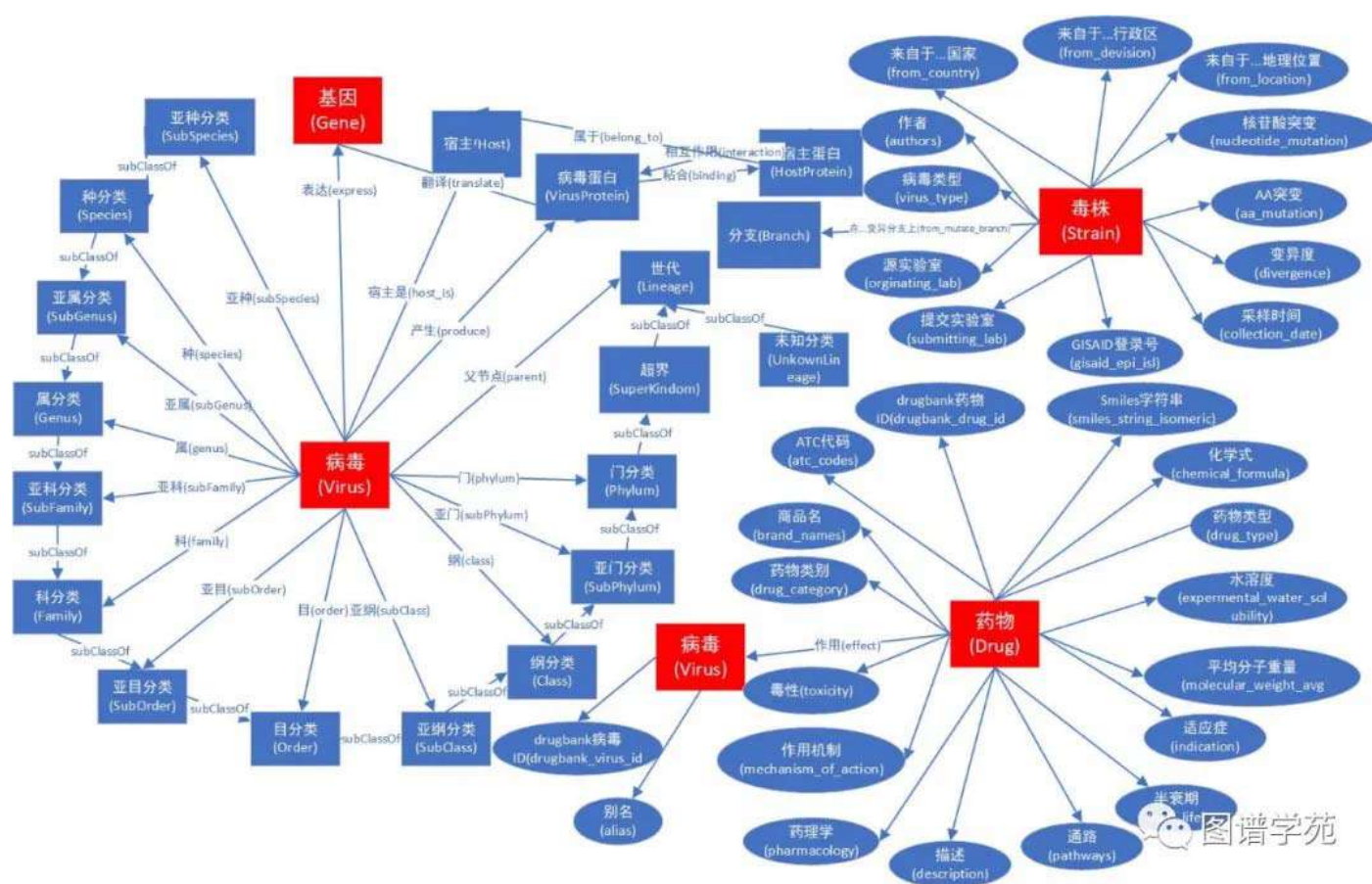
· 新冠科研图谱

同时我们还将四个子知识图谱进行了合并，合并成了新冠科研图谱，共两个知识图谱，分别为新冠科研图谱-中文、新冠科研图谱-英文。

三元组数量：2000928

实体数量：214400

谓语数量：57（谓语为四个知识图谱谓语的并集）



02 新冠百科图谱

新冠百科（包括病毒、细菌、流行病、传染病等）。一个领域百科KG，这是基础；从各大百科出发，也会涵盖之前几次疫情涉及到的知识点。

图谱以病毒, 细菌为主体, 扩展了治疗, 疾病相关内容, 通过了这些概念的百科知识, 形成新冠百科知识图谱。

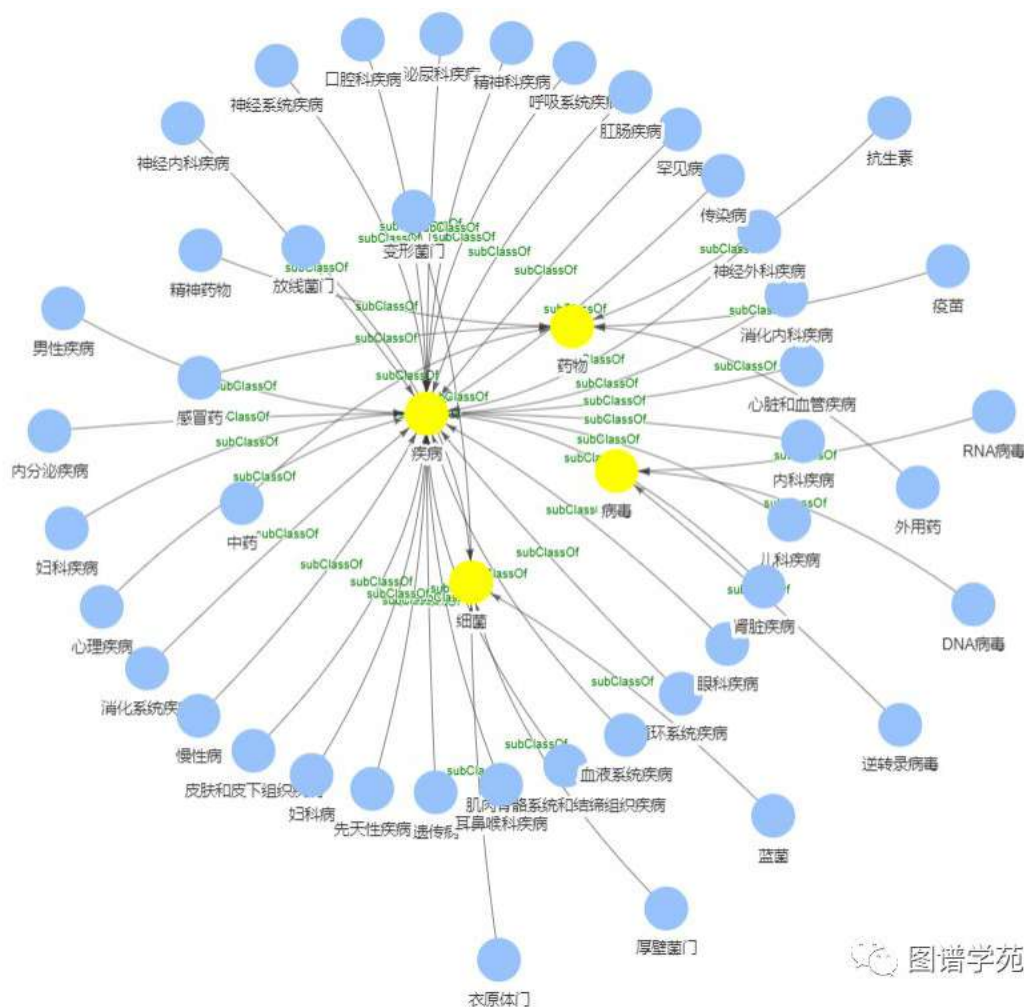
潜在应用：面向新冠相关术语的语义检索、智能问答，并可用于新冠相关文档的智能搜索和推荐。

后续工作中, 将从百度百科, 互动百科, 中文维基百科, 医学百科中进一步挖掘病毒, 细菌, 疾病, 医学之间的关系。采用基于本体的信息抽取技术, 扩充实体的属性信息。同时, 完成与zhishi.me等通用知识图谱的链接。

实体数量: 54419

关系数量: 13

谓语数量: 49



03 新冠健康图谱

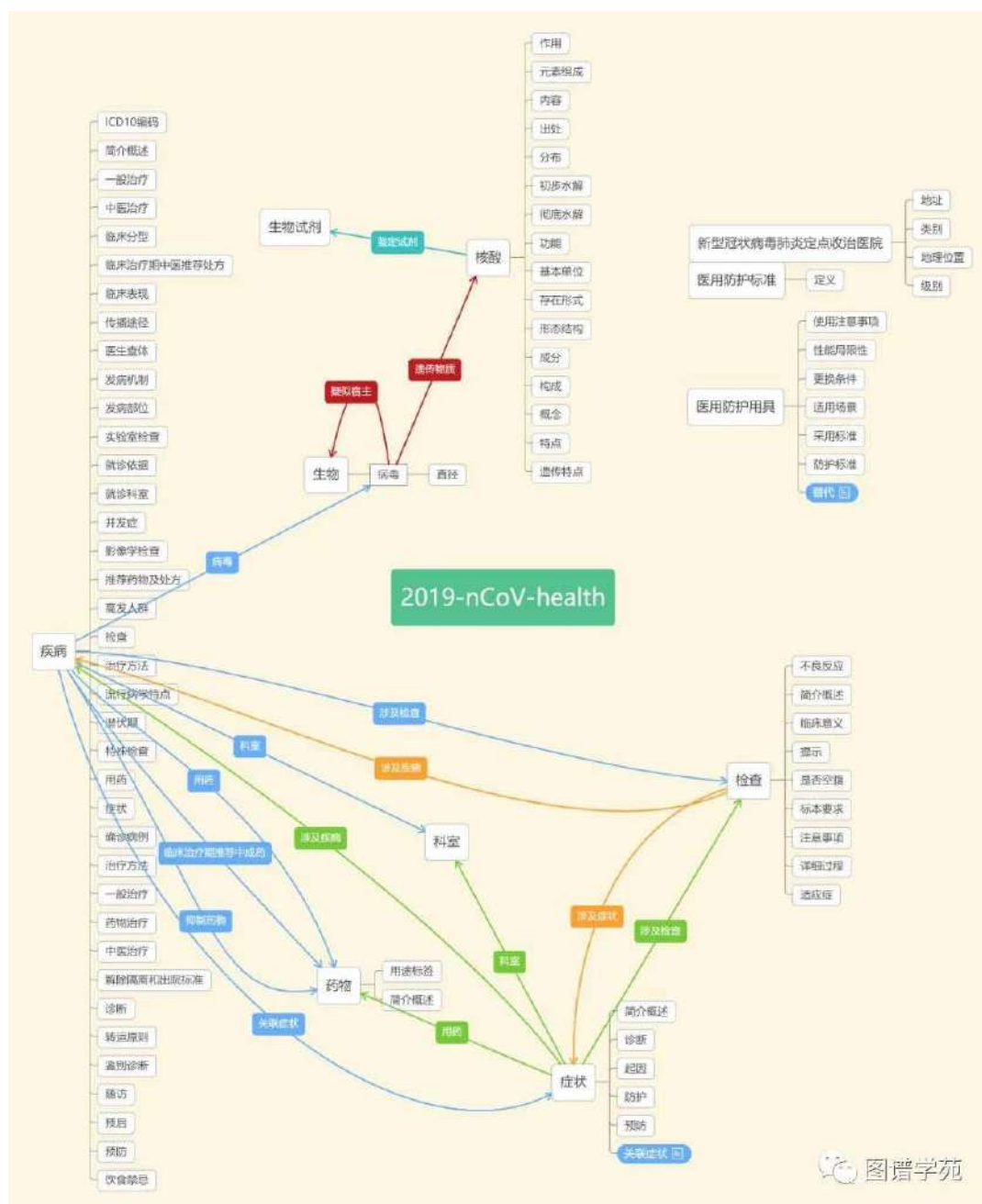
新冠健康图谱由清华大学和北京妙医佳健康科技联合构建，包含跟新冠肺炎相关的各类疾病、药物、症状、检查、全国各地接收新冠肺炎定点医院等信息。

潜在应用：流行病调查研究，基于图谱的新冠肺炎健康防护问答。

实体数量：7205

关系数量：13

谓语句数量：96



04 新冠临床图谱

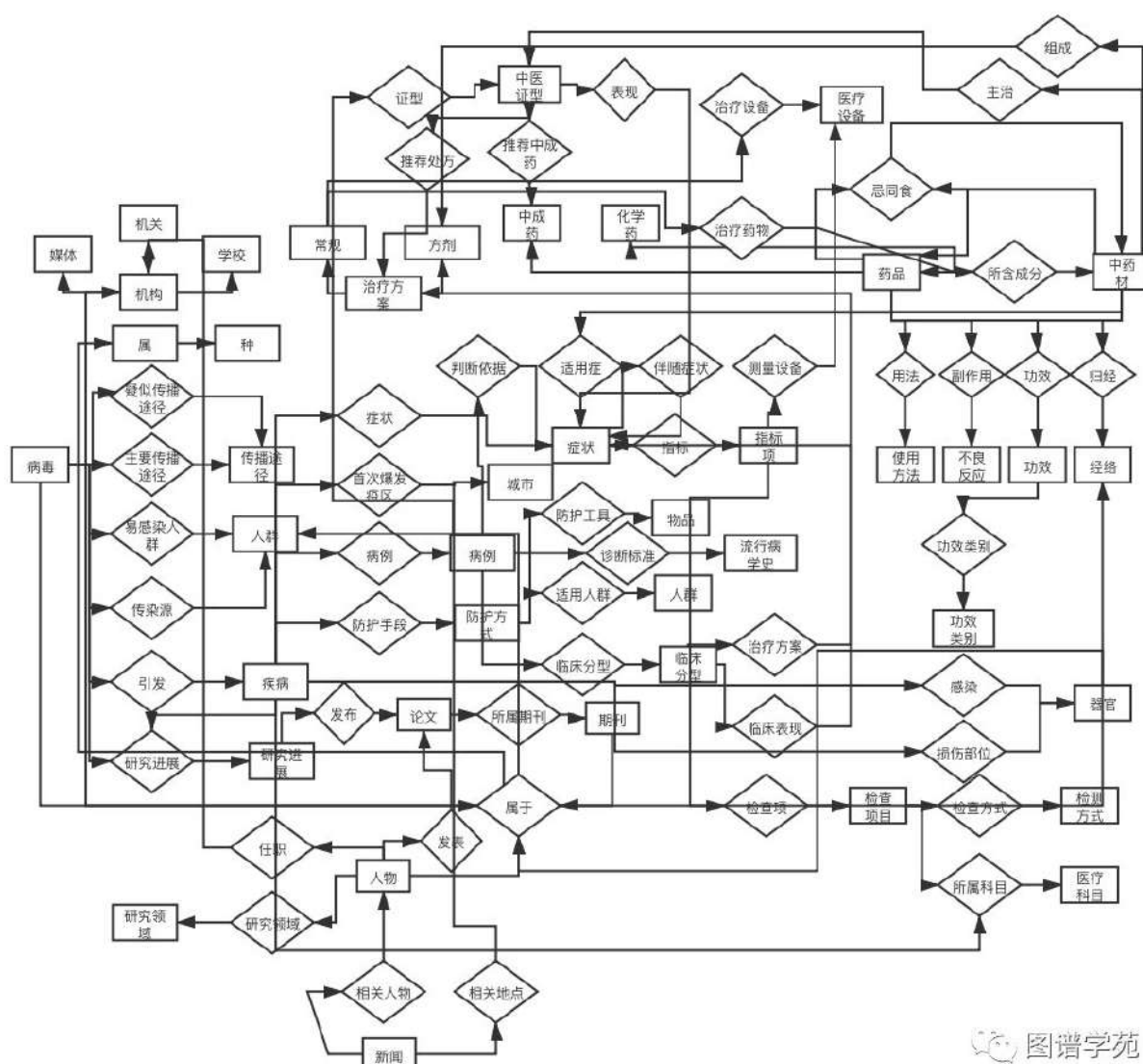
临床图谱从目前已有的规范文件入手，基于诊疗规范（流行病学+症状+实验室指标+治疗）、研究进展（测序、药物开发、疫苗）、发病统计（丁香园），将研究进展与科研相关联。

图谱以新型冠状病毒（COVID-19）为核心延展至病毒、治疗方案、症状、方剂等各类概念，形成新冠临床知识图谱，可基于图谱进行知识问答。

实体数量：383

关系数量：15

谓语数量：51



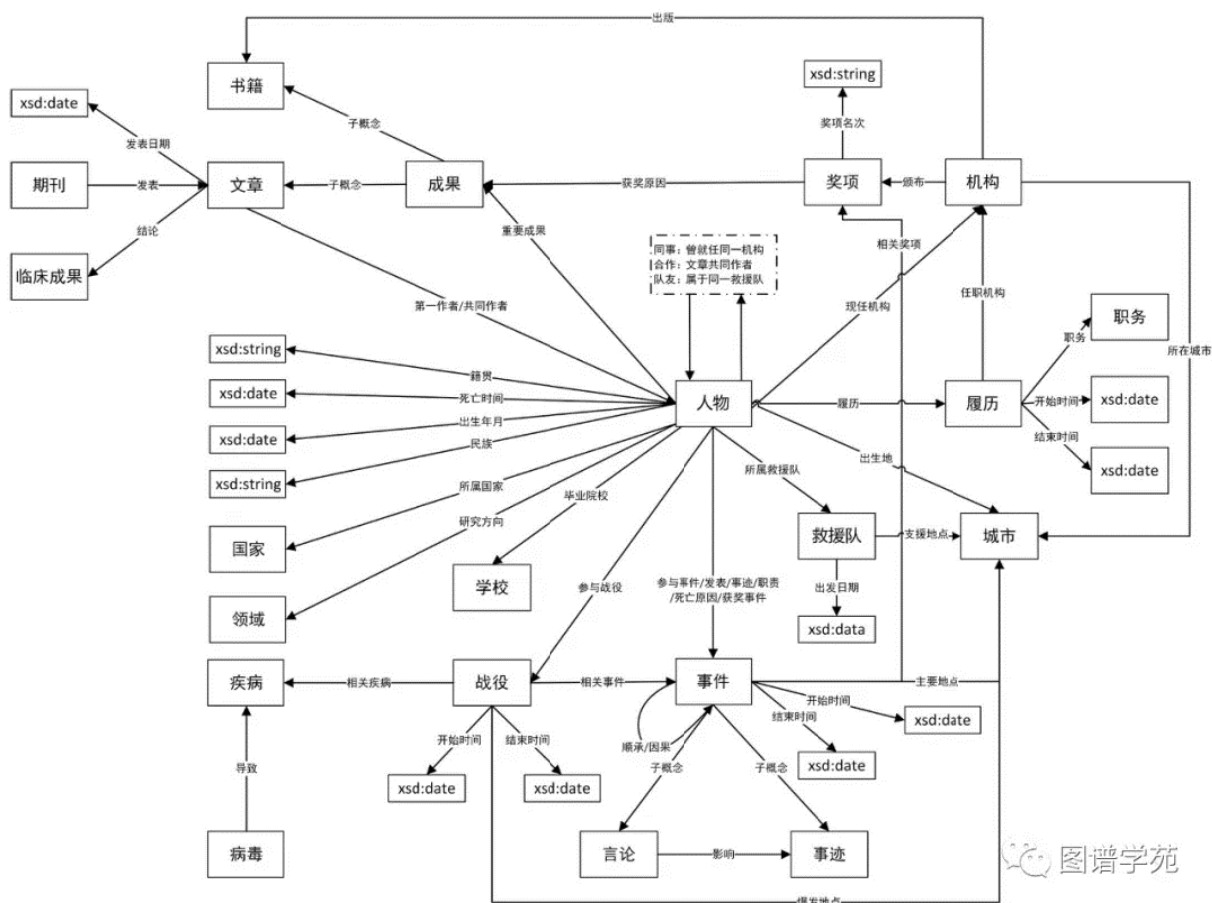
05 新冠英雄图谱

从百度百科、微信公众号、知网、澎湃新闻等获取的抗击疫情期间的部分英雄人物信息，涉及生平事迹和基本属性等，包括123个人物，其中有5位专家人员，27位因公殉职人员，91位医疗队代表人员。图谱以新冠病毒专家为核心延展至履历、成果、事件、战役等各类概念。

三元组数量：3173个

实体数量：1563个

谓语数量：46



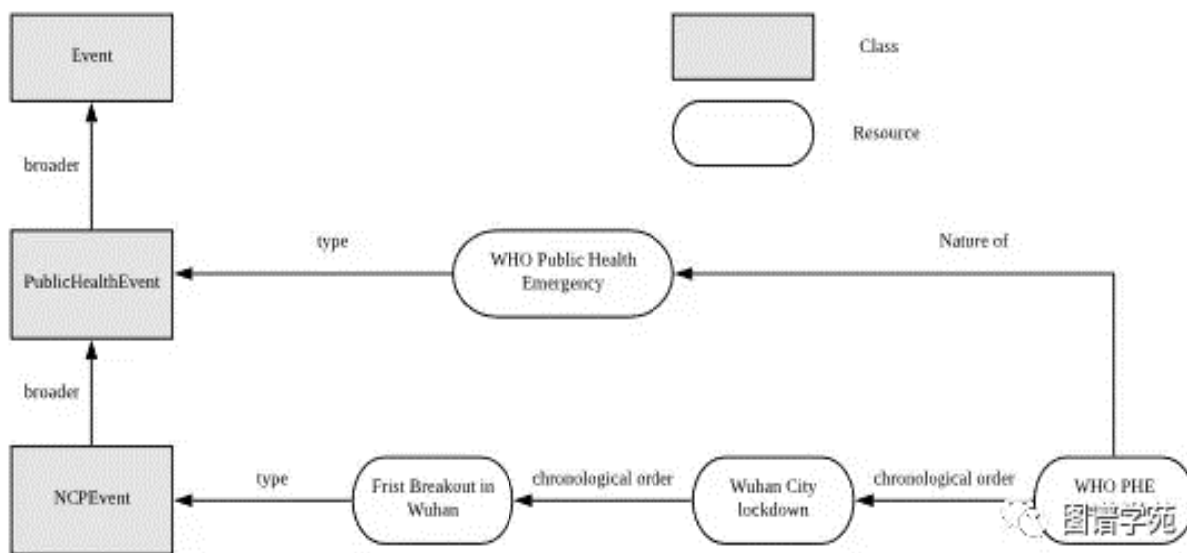
06 新冠热点事件图谱

从人民日报、丁香医生、腾讯、新浪微博抽取的新闻数据，从第一例出发到封城到各地响应等的重大事件脉络及相关时间，与新冠百科、新冠科研、新冠临床、新冠防控、新冠英雄等有所关联，并对新闻中的一些内容进行相关语义标注。支持事件在时间上的正向和反向索引，并提供事件本身发展脉络的枚举。

三元组数量：2262个

实体数量：224个

谓语数量：17个



07 新冠流行病学图谱

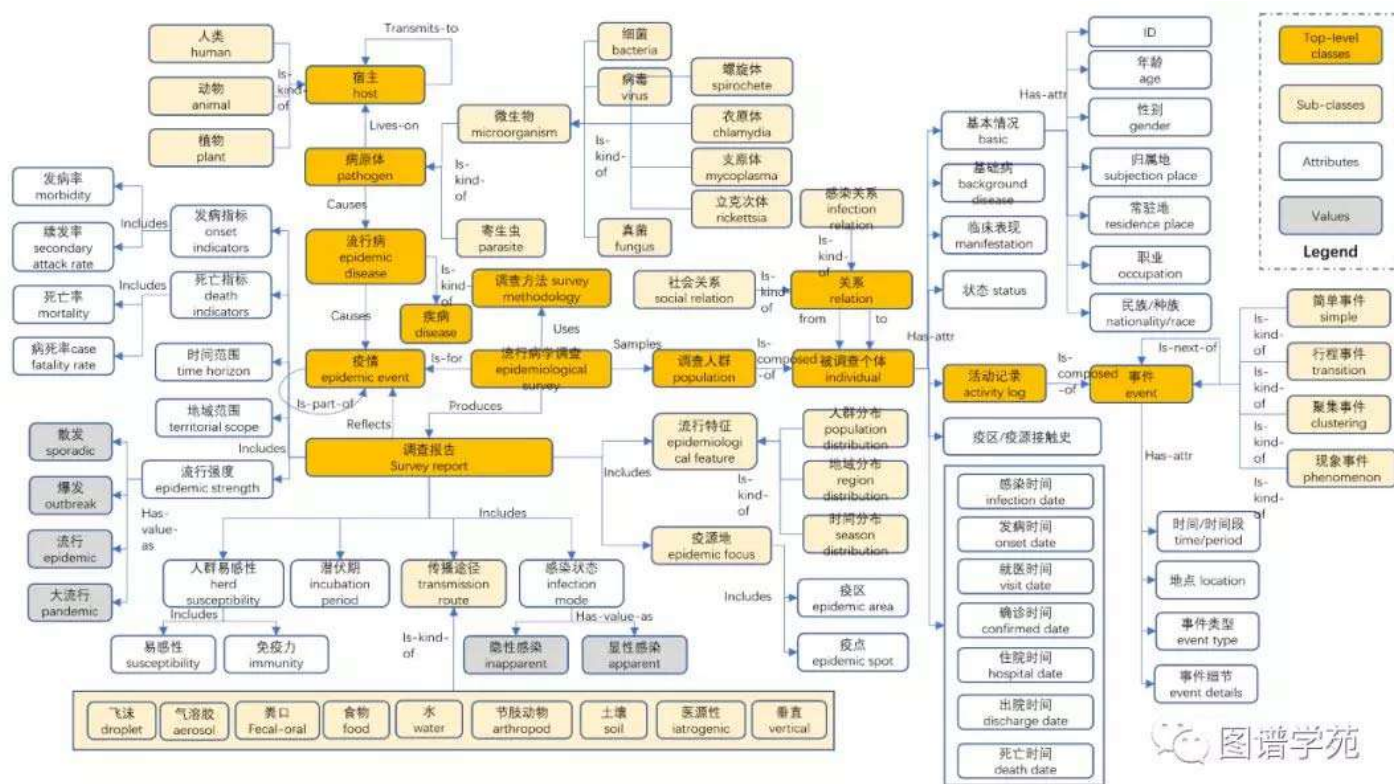
新冠肺炎流行病学知识图谱包含了流行病学知识图谱Schema和基于此Schema定义的新冠肺炎资源实例。其中流行病学知识图谱Schema重点刻画流行病学的基本概念、流行病学调查等内容，未包含“流行病学研究方法”、“预防与控制策略”、“临床治疗与预后”等内容。

目前，实例数据尚且不足。

三元组数量：374个

实体数量：240个

谓语数量：10



08 新冠物资图谱

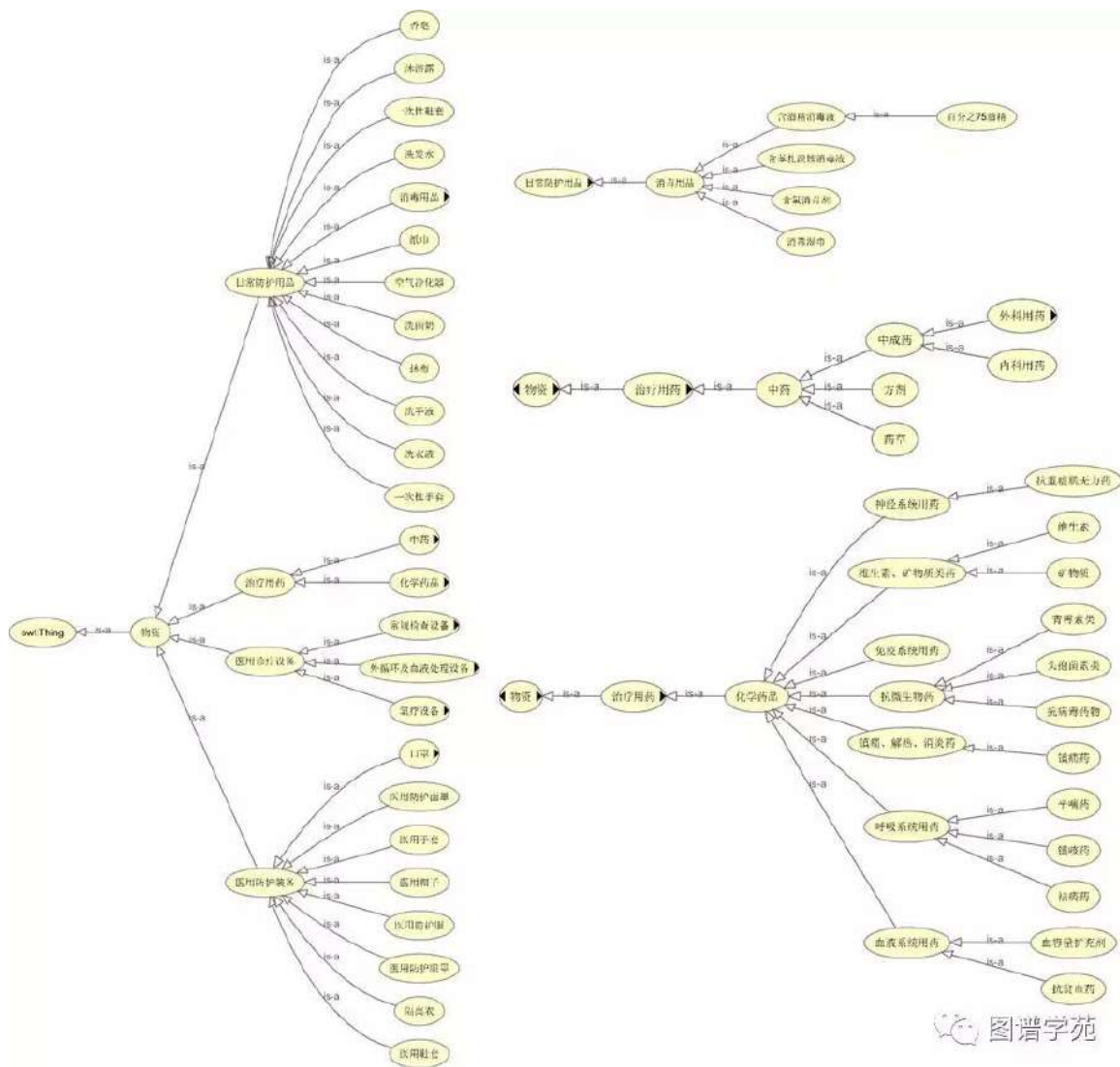
新冠物资图谱包含防控新型冠状病毒所需要的各类物资，主要包括医用防护装备、日常防护用品、医用诊疗设备以及治疗用药。其中，医用防护装备包含了医用手套、口罩、隔离服等；日常防护用品包含洗手液、消毒用品等；医用诊疗设备包含常规检查设备和氧疗设备；治疗用药包含了《新型冠状病毒感染的肺炎诊疗方案》多个版本中的中、西药品。

实例数据主要涉及部分物资的用途以及相关信息，比如：品牌，产地，使用注意事项以及保存方法等等。

三元组数量：3634个

实体数量：364个

谓语数量：43



09 结束语

本次发布的知识图谱由北京大学王选计算机研究所数据管理实验室邹磊教授团队和武汉大学信息管理学院洪亮老师团队联合构建，为了方便用户使用，我们制作了详细的图谱介绍及使用文档，您只需搜索“图谱学苑”公众号后，在文字输入框中输入“疫情图谱”即可获取相关文档链接。

（本期图说天地内容转载自微信公众号“图谱学苑”）

（本栏目责任编辑：靳丹 赵勇）