**1.新冠百科图谱**

**1.1简介以及潜在应用**

新冠百科(包括病毒、细菌、流行病、传染病等)。针对于COVID-19领域的百科KG,是所有新冠图谱的基础；从各大百科出发挖掘数据，涵盖疫情涉及到的相关知识。

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

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。

3.0版本从英文维基百科页面出发，完成了英文图谱构建，并实现了中英文跨语言链接。

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。

**1.2数据schema**

详见文件baike-schema.xlsx和baike-schema\_en.xlsx

**1.3数据来源与规模**

**数据来源**

当前数据从百科抽取，以infobox作为主要数据源；

百度百科：通过类别获取数据

互动百科：通过类别获取数据

中文维基百科：通过类别获取数据

医学百科：通过类别获取数据

英文维基百科：通过类别获取数据

目前选择的实体类别限定于：病毒，疾病，细菌，症状，医学专科，检查科目，药物

**规模**

百度百科：实例30390个，三元组106264‬个

互动百科：实例38,310个，三元组102360个

中文维基百科：实例1695个，三元组2144个

医学百科：实例26852个，三元组31031个

英文百科：实例11051个，三元组56864个

**1.4规范**

**命名空间**

概念 <http://www.openkg.cn/COVID-19/wiki/class/C> + ID

属性 <http://www.openkg.cn/COVID-19/wiki/property/P> + ID

实体 <http://www.openkg.cn/COVID-19/wiki/resource/R> + ID

**1.5作者介绍**

李志强 zqli@seu.edu.com 东南大学苏州联合研究生院

秦旭 qinxu126@163.com 东南大学苏州联合研究生院

许茜 xuxi0127@126.com 东南大学苏州联合研究生院

李林 li\_lin@seu.edu.cn 东南大学计算机学院

殷春锁 213180808@seu.edu.cn 东南大学计算机学院

吴亦珂 wuyike5074519@163.com 东南大学人工智能学院

王然 2965057336@qq.com 东南大学软件学院

吴天星 wutianxing@seu.edu.cn 东南大学计算机学院

**2.6可视化样例**

