

知识图谱定义及可视化系统

COIN: A system for CONstructing and vsualizing kNOWLEDge graph

知识图谱（Knowledge Graph）的概念由谷歌 2012 年正式提出，旨在实现更智能的搜索引擎，并且于 2013 年以后开始在学术界和业界普及。目前，随着智能信息服务应用的不断发展，知识图谱已被广泛应用于智能搜索、智能问答、个性化推荐、情报分析、反欺诈等领域。知识图谱以其强大的语义处理能力与开放互联能力，可为万维网上的知识互联奠定扎实的基础，使 Web 3.0 提出的“知识之网”愿景成为了可能。知识图谱是一种结构化的语义知识库，用于迅速描述物理世界中的概念及其相互关系。通常知识图谱通过对错综复杂的文档的数据进行有效的加工、处理、整合，转化为简单、清晰的“实体,关系,实体”的三元组，最后聚合大量知识，从而实现知识的快速响应和推理。

本项目旨在构建一个系统平台通过在线编辑基本图元以可视化的方式构建个结构完整、业务覆盖全面的知识图谱知识图谱，并初步具备一定知识推理能力。

迭代一：

1. 识别本体、关系等关键要素

初步识别知识图谱可视化相关的要素，如实体、关系等；

2. 定义关键要素图元

初步定义实体、关系等的图形表示属性，如颜色、虚实线等；

3. 定义知识图谱的基本结构

定义或者通过 Schema 文件等描述知识知识图谱的合法结构

4. 可视化在线图谱

能够读取预先定义的知识图谱，并进行在线显示；

5. 可以通过在线编辑图谱

能够在线以文本的形式编辑知识图谱；

6. 知识图谱导出

能够将可视化的只是图谱导出 xml，图片等文件；

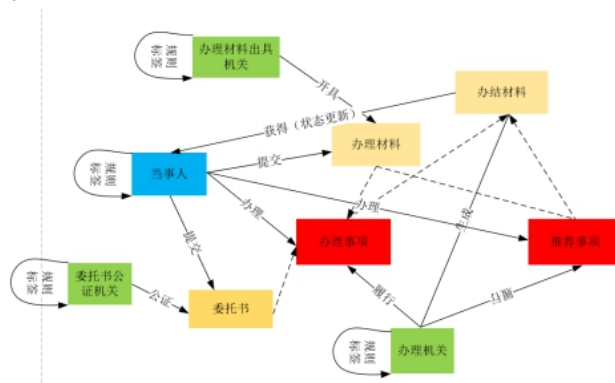
7. 能够复现下述实例

初步实现以下示例的读入、编辑和保存。

构建数据图谱本体模型的方法，是从政策文件中，提炼出政策条例中的本体模型相关的词汇，以关系网络的方式进行组织，将要素定义为实体，将要素之间的关系定义为边。以“收养子女登记”为例，提炼部分的词汇与要素方法如下：

作为一个“政务事项”，这个事项在具体的办理过程中，称为“事项”。事项办理的申请人为“当事人”。事项办理由相应的政府部门来执行，称为“办理机关”。办理过程需要提交相关的“办理材料”。这些材料的类型如证明、协议、证件、文书、申请书等。

举例图示如下所示：



上述举例说明的本体模型要素的梳理，只是作为本体定义的一个思路 and 说明。上图仅供参考，本体关系保留，图元表现形式请自行定义。

检查点：

1. 能够读入预定义的简单知识图谱。
 - 不统一给定预定义格式
 - 每组给出自己的测试数
 - 具备简单的实体关系即可
2. 能够在线编辑知识图谱
 - 可以在线增加、删除、修改图谱中的实体
 - 可以在线增加、删除、修改图谱中的关系
3. 能够持久化图谱
 - 可以导出为数据文件
 - 可以导出为图片

通过视频录制，逐一解释上述功能，视频时长不超过 5 分钟。

迭代二：

1. 完善关键要素、图元的定义
 - 保证该系统能够构建通用型数据库；
2. 允许用户自定义图元

允许用户添加图元，定义或修改图元属性；

3. 支持在线图像化编辑图谱

能够在线通过拖拽等方式编辑图谱

4. 能够自适应展示布局

默认情况下，能够以较优的方式展示图谱；

5. 能够在线编辑图谱布局

可以在线通过拖拽的方式调整图谱布局。

6. 能够保存所调整图谱布局

编辑过布局的图谱再次读入时按编辑时布局展示

7. 使用数据库

使用 Mysql 等数据库保存系统数据

8. 具备但不限于以下可视化能力

- ✓ 节点搜索：搜索图谱内容，包括节点、关系，节点属性的取值。支持模糊匹配。支持搜索历史操作的提示，支持对命中的节点进行标记或者高亮。
- ✓ 类型过滤：可以根据节点类型对图谱进行过滤，从而简化图谱信息。
- ✓ 展示效果调节：调整关系网络图的显示效果，包括调整节点之间的距离、节点图标直径大小、调整节点文字大小、是否显示关系的标签。
- ✓ 图谱统计：显示当前打开的图谱的统计数据。
- ✓ 缩放：画布区域图的缩放功能，设置按键键可退回到初始打开时的位置。
- ✓ 显示模式切换：包括排版模式和力导图模式，两类模式相互独立，互不影响。排版模式是将实体和关系进行分类，类似于阅兵的方阵，要保障良好的视觉效果。支持排版模式下实体和关系的位置的持久化。

检查点：

1. 能够自定义图元

可以增加、删除、修改图元

涵盖图元的图形、属性

2. 能够调整及保存图谱布局

可以在线编辑图谱布局

可以保存、读取图谱布局信息

3. 满足 8 中要求的可视化能力

通过视频录制，逐一解释上述功能，视频时长不超过 15 分钟。