

## 毕业设计软硬件验收表

任务书规定完成的软硬件内容及规定的技术指标：（学生填写）

上传的附件中是我实现的“基于多知识库的表格实体链接系统”。

其中包括了 2 个实体链接的方法，方法一是我在 JIST 2016 论文 (Entity Linking in Web Tables with Multiple Linked Knowledge Bases) 中的方法；方法二是毕设过程中新提出的方法。

2 个方法都是基于图模型的随机游走算法。

系统接收的输入为：

1. 表格
2. 知识库各类型数据 (abstracts, labels, infobox\_propertyies, sameAs)
3. 表格的人工标注数据 (用于评估链接效果)
4. BabelNet 的同义词数据

有了这些数据，将他们进行一些格式上的规范化，在系统中修改数据文件的路径，算法就可以成功的运行了！

系统的输出有：

1. 表格中 mention 的候选实体集合文件
2. 方法一单库链接的结果 (每张表格的链接结果，每个表格的实体消歧图)
3. 方法一多库链接的结果 (每张表格的链接结果，每个表格的实体消歧图)
4. 方法二多库链接的结果 (每张表格的链接结果，每个表格的实体消歧图)
5. 一个包含所有表格的链接结果的文件，算法运行完成后会将所有表格的链接结果合并到一个文件中
6. 每张表格链接的准确率、召回率、F1 值和 MRR 值，在所有表格处理完后，会输出所有表格综合的链接准确率、召回率和 F1 值
7. 运行日志 (log)

生成表格中 Mention 的候选实体过程耗时较长，因为知识库中的实体数量规模巨大。

有了候选实体数据之后，运行实体消歧算法并得到最终的消歧结果大约平均 1 分钟/每张表。这 1 分钟里大部分时间是花费在了判断来自 2 个不同知识库的实体有没有 sameAs 关系上，除去这个部分，每张表消歧其实只需要大约 5 秒左右。

学生签名：严晟嘉

2017 年 5 月 20 日

验收结果：（验收负责人填写）

☐ 超额完成任务书指标要求

☐ 符合任务书指标要求

☐ 基本符合任务书指标要求

☐ 不符合任务书指标要求

验收小组成员签字：

年 月 日

注：软件指计算机软件、图纸等。

硬件指机电装置、电路板、新材料、新制剂、结构模型等物质成果。