

毕业设计选题申报表

班级：090131

指导教师：李慧颖

<p>1、课题内容（名称、具体任务）</p> <p>名称：基于多知识库的表格实体链接研究</p> <p>任务：万维网中包含了大规模的关系型数据，它们很多都存在于 HTML 表格中。为了从这些表格中抽取有价值的信息，关键的一步就是实体链接技术。目前数据的规模已经达到了过载的程度，这导致人们在找到有用的信息之前，需要处理很多无用信息，而不能高效地找到有用的信息的原因就在于自然语言表达的多样性和歧义性。该课题的任务就是实现一个基于多知识库的表格实体链接系统来解决上述问题。</p> <p>技术路线：</p> <ol style="list-style-type: none">1) 实体链接系统需求分析2) 爬取表格数据3) 数据预处理4) 候选实体生成5) 实体消除歧义6) 系统测试与评估
<p>2、课题性质（毕业设计、科研论文），课题来源（生产、科研、自拟）</p> <p>课题类别：毕业设计</p> <p>课题来源：科研</p> <p>是否卓越课题：否</p>
<p>3、课题难易程度及作为毕业设计（论文）选题的可行性</p> <p>该课题旨在提出一种统一的架构以解决表格数据中的实体链接任务，并且期望效果要优于或至少持平于目前最先进的实体链接技术。现有的工作都是基于单一知识库进行实体链接，而实际上任何知识库都是不完整的，我们的工具期望支持多种知识库进行实体链接，从而缓解知识库的不完整的问题。基于现有的实体链接技术的研究，该课题可行。</p>
<p>4、已具备的条件（经费、设备及指导人力）和存在的困难</p>
<p>5、教学要求与工作计划</p> <ol style="list-style-type: none">1) 网络爬虫技术2) 机器学习算法3) 信息检索算法4) JAVA 或 C++或 python 编程5) 知识图谱相关技术 <p>工作计划：</p> <p>2017.1：基础知识的学习及完备</p> <p>2017.2：分析需求与初期建模</p> <p>2017.3：爬取数据与模型实现</p> <p>2017.4：系统研发与评测</p>

2017.5-2017.6：毕业论文撰写			
6、系教学主任审核意见 同意			
7、选择本课题的同学：			
姓 名	学 号	姓 名	学 号
严晟嘉	09013119		
8、院(系)教学院长 (主任) 意见			

注：

- 1、 指导教师填写上述 1～5 项。学生填上述第 7 项。
- 2、 一个学生一个题目。需要多个学生完成的选题，分别填报选题。
- 3、 下列情况的题目不宜安排学生做毕业设计（论文）：

（1）课题偏离本专业所学基本知识；（2）课题范围过专过窄，达不到全面训练的目的；（3）属本科生难以胜任的高新技术；（4）毕业设计期间难以完成或不能取得阶段成果的。
- 4、 每个同学填报 3 个志愿。为了使同学的志愿尽量得到满足，同学的志愿应有一定的分布，选择某个课题的同学不应太集中。若填报某项课题的同学太集中，人数超过了需要人数，可能会被调剂。