

2016 年秋季学期总结

Zhe, Han

Institute of Computer Science & Technology
Peking University

2017 年 1 月 13 日

- 找工作
- 学术
- 科研 && 实验室工作

- 杭州蚂蚁金服
- 9.10 拿到 offer, 后面找工作的时候划水...

- 终于发了文章
 - Detecting Synonymous Predicates from Online Encyclopedia with Rich Features
 - AIRS 2016 长文
 - 11.24 骑 ofo 去清华做了报告
- 一直拖着的知识库构建的期刊...

● 知识库

- 完成了[百度百科](#)数据爬取、整理
 - 各种坑：页面存储、磁盘空间、页面去重、别名表抽取与排序、PKUBase 类别对应
 - 1100w 实体,4500w 三元组,1800w 的类别信息,1500w 的别名表
- 找到了接班人...
 - 带吴雨婷师妹做了一些东西（任务很艰巨），希望把知识库的维护做下去
- 做了[新版的前端](#)展示界面
 - 自认为界面友好度 or 美观程度提高了不少...
- 方正提供了一个知识库版本

● 新华社新闻分类

- 回来后整理了一下之前的数据
- 帮老师跑了几次会、交了一些材料，终于算是结项了

- 下一步工作

- 进行中：知识库**实体链接**
- 进行中：百度百科与维基百科的融合
 - 目前 2 个知识库版本已经有了各自的别名表、三元组、PKUBase 类别信息
- **期刊...**

谢谢大家！