

# 第一次小组会议5.17

现阶段已完成的知识抽取非常基本，只是用pyltp完成了对文本的主谓宾切分，构成的三元组并不是知识的三元组。

下一步我们需要完成知识库的构建、抽取出真正的知识

知识库领域现在初步定在金融与医学领域。

知识融合：<https://blog.csdn.net/yangfengling1023/article/details/82117997>

知识图谱构建技术概要：[https://blog.csdn.net/lairongxuan/article/details/100527179?utm\\_medium=distribute.pc\\_relevant.none-task-blog-2~default~BlogCommendFromMachineLearnPai2~default-1.control&depth\\_1-utm\\_source=distribute.pc\\_relevant.none-task-blog-2~default~BlogCommendFromMachineLearnPai2~default-1.control](https://blog.csdn.net/lairongxuan/article/details/100527179?utm_medium=distribute.pc_relevant.none-task-blog-2~default~BlogCommendFromMachineLearnPai2~default-1.control&depth_1-utm_source=distribute.pc_relevant.none-task-blog-2~default~BlogCommendFromMachineLearnPai2~default-1.control)

第一周任务：

研究pyltp的具体作用？

基于文本提取出实体和实体之间的关系（如果要加知识库、语料库，那就是金融和医学领域）

## 3. 知识图谱关系抽取

仅用于测试用，可以pip3 install jiagu==0.1.8，只能使用百科的描述进行测试。效果更佳的后期将会开放api。

```
import jiagu

# 吻别是由张学友演唱的一首歌曲。
# 《盗墓笔记》是2014年欢瑞世纪影视传媒股份有限公司出品的一部网络季播剧，改编自南派三叔所著的同名小说，由郑保瑞和罗

text = '姚明1980年9月12日出生于上海市徐汇区，祖籍江苏省苏州市吴江区震泽镇，前中国职业篮球运动员，司职中锋，现任中
knowledge = jiagu.knowledge(text)
print(knowledge)
```

训练数据：<https://github.com/ownthink/KnowledgeGraphData>

实体的类型主要包括三大类七小类：

- 实体类（包括人名，地名，机构名）
- 时间类（日期，时间）
- 数字类（货币、百分比）

第二周任务：

基于不同的数据源（语料库）完成关系抽取、属性抽取

第三周任务：

知识融合：主要关键技术包含指代消解、实体消歧、实体链接。

