

# CN-Dbpedia阅读报告

---

文章核心是以下几部分工作

- 动机：以往知识构建过程总有大量人工参与，作者想要减少人工参与的工作，所以直接用了DBpedia的本体，但是需要做的是将中文实体分类到相应的类别中去
- 第一部分是减少本体构建的人工参与，作者直接用DBpedia的本体，那么主要工作是将中文实体分类到英文类别中，作者构建训练集的方法是找到同样中文标签下的中英文实体，然后根据英文实体的类别，构造出训练集。
- 动机：实体infobox的信息大都是缺失状态，需要补充。维基页面中含有丰富的信息(虽然这里作者没有提出，个人认为，作者只是利用了维基页面中的摘要)
- 第二部分是去补全infobox的信息，作者为每一个关系制作了一个抽取的模型，是一个类似于seq2seq的lstm模型，作者将维基百科页面输入进去，输出也是一个等长的序列，标记每一个词是不是待抽取的结果。例如：

...DiCaprio was born in Hollywood , California , the only child of ...

0 0 0 0 1 0 1 0 0 0 0 0

- 动机：以前的更新策略，是一阶段更新一次，但是问题就是多久更新一次，时间长了，可能部分信息都过时了。时间短了，对于带宽和计算资源的压力过大。所以作者提出热更新这个概念，对于改变状态的实体进行更新。
- 第三部分工作是作者提出了一种更新策略，相比于以前的阶段性的更新策略，作者利用新闻中涉及到的实体去更新实体，另一方面是从搜索引擎的热词搜索来得到实体。
- 这其中有一个疑问，作者并没有把更新策略讲清楚，是将整个实体更新，还是实体的某个属性更新，更新方法同第二点一样吗