

This the note for : 2013 Mining Evidences for Named Entity Disambiguation

问题

希望将未知实体（包括它的一段**短文**，比如40个词）对齐到知识库中的命名实体。如果知识库中对应实体对应的文档与短文不匹配，可能无法准确对齐。

解决方案

作者引入了外部的信息(external evidence，对应的知识库中的信息叫做 internal evidence)来辅助对齐。知识库对一个实体的描述是一个文档和若干超链接。

利用外部信息的方法是，从google搜索中检索前20个页面，然后将这20个页面的文档（长文本）与知识库中的文档做对齐（长文本对齐是相对容易的）。相当于又标准了一批文档。

具体的，对齐被看作一个分类问题，将待定实体划分为某一个已有的实体。作者采用了一个生成模型来描述有一个实体来生成一个文档的过程。

