金融系统

1.输入数据的形式是什么？

有标题、来源、正文等这种数据，就想excel中的表格那样

1. 公司中名称和简称的识别【主体识别】

输入：公司的实体

输出：对应这篇新闻中的主体公司

需要输出所有公司吗？ 还是只需要输出新闻中的主体公司？

**需要输出一篇文章中的所有公司**

文章中某个公司出现的次数越高，是否就代表这个文章跟这个主体公司有关？

1. 新闻主体的多标签识别

新闻的表

输出：标签（前面输入的主体公司的标签）

1. 整个系统的输入是文章。确认是否做了结构化？（已经结构化了，有标题，有时间）
2. 输出的是数字：

输出：计算该企业的营销风险指数

1. GRPC 和HTTP接口反馈
2. 数据实体标识（在本项目中只做企业相关）

问题：

01.基金组织名算不算？

02.国资委，经信委算不算？

1. 有1.2亿个体名称

成熟的公司名称

10.容易混淆的标签

01.'增资扩股', '股权增持’

02.收购，公司收购 => 能否将收购再细化？也就是说二者应该保持在一个粒度

03.批准，批准发行，批准进入…  => 标签粒度不同。能否删除批准这个标签？

04.违法行为，专业违规    => 这两者实际含义是否相同？

05.创新，科技创新 => 仍然是划分粒度的问题