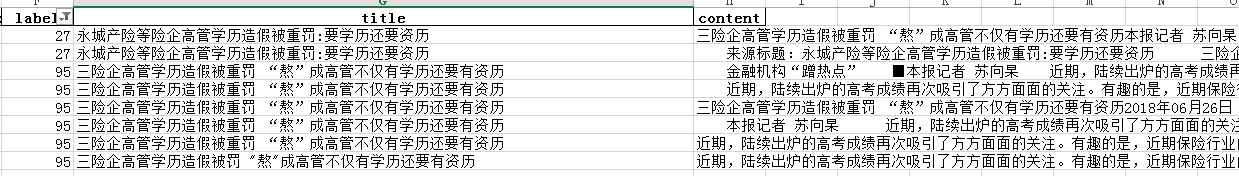
1. 区分不开：匹配到多条（test\_mix\_1）

同一事件，不同文字报道，但是业务认为是不同新闻。

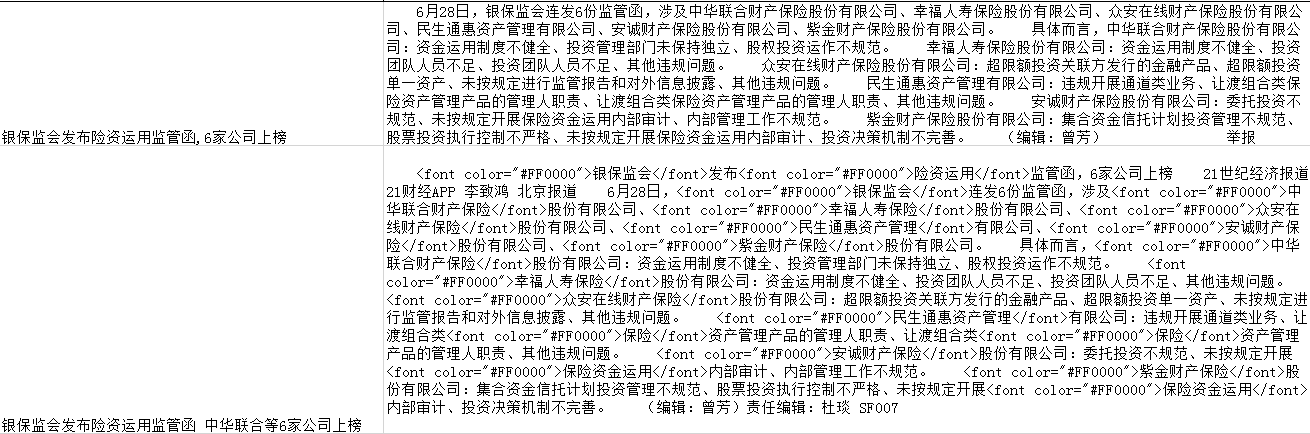






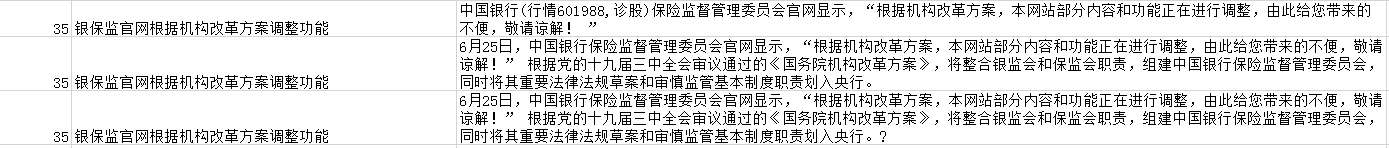
二、区分不开：匹配不到数据（test\_mix\_2）

同一事件，前面有噪音





同一事件，文字长度不一样



三、匹配错误（test\_wrong）

同一事件，不同文字报道，但是业务认为是不同新闻。

