数据处理

1. 提取合适的公司（招聘数量top1000）；
2. 对每个Post分词、去stop-words；
3. 对公司按招聘聚类，获取公司的近邻；
4. 建模；