

# 企业相关信息收集

## 企业相关信息收集

- 1. 背景
- 2. 需求
- 3. 数据示例

## 1. 背景

产业链的关键词提取和补充时，从已有的企业库中公司的经营范围、主营产品等信息中抽取出的关键词效果较差，分析原因发现现有的企业的经营范围描述文本不够准确，导致从里面抽取出的关键词跟公司实际的经营情况不一致，例如：医药公司的经营范围里有"销售化妆品", "五金产品批发"等文本。因此需要一批准确描述公司经营范围的数据，从而更好的完成关键词的提取和扩充。

## 2. 需求

待收集字段	数量	准确度	覆盖率
公司名、公司经营范围、公司主营产品	500w+	经营范围抽样查看 准确率>95%	公司名100%, 经营范围100%, 主营产品>30%

注意：已有的企业库的公司经营范围信息是从天眼查购买，描述不够准确，请勿再从天眼查等网站中收集此类信息。

## 3. 数据示例

```
{
  'name': '长春高新',
  'business_scope': '高新技术产品的开发、生产、销售及服务基础设施的开发建设；物业管理；高新成果转让及中介服务；商业供销业（国家有专项限制经营的商品除外）；培训、新药开发、技术转让，咨询服务；集中供热；产业投资（医药产业）（以上各项仅限分公司、子公司持证经营）。绿化景观工程设计、施工及维护、企业营销策划、广告策划、社会经济信息咨询、房地产中介服务、房屋销售代理（仅限分支机构持证经营）；企业管理咨询（以上各项国家法律法规禁止的不得经营；应经专项审批的项目未获得批准之前不得经营）',
  'products': ['重组人生长激素', '重组人促卵泡激素', '水痘减毒活疫苗', '人用狂犬疫苗', '鼻喷流感减毒活疫苗', '血栓心脉宁片', '银花泌炎灵片']
}
```