1. 相似企业查找
2. **专利检索模块**：给定企业，在线从互联网上（soopat 、http://www.sipo.gov.cn/或者其他数据源）检索该企业的所有发明专利的基本信息和摘要，包括发明专利名称、申请人、专利日期、申请号、状态和摘要。
3. **专利信息抽取及企业专利核心关键词选取模块**：抽取每个专利的专利名称和专利摘要中抽取关键词关键字， 然后从所有检索到的专利中选择权重最高的1到N个关键词作为给企业专利信息的核心关键词（可能提取的是短句或者拼接的短句）。
4. 关键词测试：对提取的核心关键词进行模拟测试，查询关键词对应的发明专利数量，若数量很大（数量大于N），且其中X%以上的专利所有者为规模很大的企业（这些企业需要前期测试给出列表，比如专利所有者最多的前一百企业），则说明这个关键词比较普通，效果不好。
5. **关键词相关专利及相关企业检索模块**：通过上一步的1到N个关键词检索发明专利，并抽取检索结果中的发明专利名称、申请人、专利日期、申请号、状态和摘要，并统计其中的唯一申请人（企业）形成相关企业集合。仅提取申请人为企业的，个人、科研机构、国外机构均剔除。如果申请人中既包含企业又包含个人或科研机构，则保留。（企业指名称末四位为【有限公司】）。
6. **相关企业的专利检索模块**：对检索的每个企业，检索其专利，记录专利个数。
7. **数据输出模块**：对检索到的企业及专利进行排序（按照企业名称、专利数目等对企业进行排序），并导出列表到Excel。
8. 企业基本信息爬取
9. **企业信息检索功能**：到企查查或类似网站检索企业基本信息包括注册资本、股东、股权变动、联系方式、企业类型等（具体内容待定）。剔除企业类型为上市的。
10. 针对企业股东中的法人股东，还需要查询证券投资基金业协会，查询该股东所属的管理公司，并爬取管理公司名称及备案信息的链接

<http://gs.amac.org.cn/amac-infodisc/res/pof/fund/index.html> 基金公示

1. **企业基本信息和资讯检索模块**：利用上述（1）和（2），检索企业基本信息和相关资讯。资讯检索使用“企业简称+上市”，“企业简称+融资”，“企业简称+投资”“企业简称+外资”，“企业简称+国有”等，具体使用的关键词待定。提取资讯中包含关键词的那部分信息。
2. 输出爬取后的企业信息。1和2连接到一起自动进行，最终输出的是一个excel表格，其中企业是通过专利法得到的，后面对应的是企业的专利数和基本信息。