**第一阶段：打通所有采集解析基本功能**

1. mongoDB存储模型建立，转换解析结果由map类型到mongo存储类型，实现解析数据在mongoDB的存储，形成原始情报库。
2. 实现mongoDB的统一dao方法，实现解析功能的增量爬取，每次解析删除mongoDB中旧的内容，刷新存入新的内容。
3. 博客，新媒体（豆瓣等）等站点的爬取，解析功能类编写（尚不完善）。

**第二阶段：预检查机制和日志系统的设计与实现**

1. 预检查功能编写，实现机制类似于工具包中其他的功能，采用反射机制，并打包成统一的检查执行类，根据后期任务日志的任务状态不同，开启任务时，检查项目不同。
2. 任务日志三层表结构及相互逻辑关系设计，设计数据库表。
3. 向爬虫内核植入日志系统，实现三层日志表结构的实时数据刷新和日志写入功能（实时统计采集器运行结果，写入采集器各级日志）。

**第三阶段：多线程并行任务实现**

1. 实现多站点并行处理采集任务的多线程队列功能实现，并整合入总任务日志的创建和状态实时更改。
2. 实现总线程暂停功能实现，并记录当前采集状态，实现恢复任务后从上次未完成的阶段开始任务，不需重新开启任务的功能。
3. 总任务暂停，终止，取消，恢复，开启，等功能方法的编写，可直接传入前台action调用。

**第四阶段：配合冯师兄将采集器功能整合入总项目**

1. 配合冯师兄将采集器所有功能整合入总项目（其实都是冯师兄做的）。
2. 完成采集器功能在总项目中的基本测试。
3. 完成采集器整合后在服务器上的初步压力测试，运行正常。

**后期工作规划**

1. 整合lzr B页面各种按钮的action逻辑（包括暂停，终止，取消，开启和恢复，由于逻辑功能涉及到线程调度的特性，需要本人编写）。
2. 站点数量大规模增加。
3. 完善博客，新媒体（豆瓣等）的爬取，解析所有功能。
4. 采集器在总项目上完美运行，大规模压力测试。
5. \*附加功能1，爬取，解析参数的更新修改功能增强：parame参数逐个手动修改太麻烦了，对用户体验不佳，需要采取升级包机制：维护人员提供升级xml配置文件，升级上传xml配置文件，自动修改parame参数数据库，完成升级。（需要的技术功能：文件上传，上传文件的读取，xml配置文件的解析，根据解析结果对表修改）。

\*附加功能2，新浪微博模拟登陆功能尽可能实现，微信公众平台尝试。