**我的爬虫V1.0设计方案**

设计目的：

1. 本设计主要用于日常网络搜集数据需要；
2. 本设计拟采用通用架构设计模式，以练手的方式强化设计模式的学习；
3. 加强MongoDb和Qt的使用；
4. 加强NumPy、Pandas和Matplotlib的使用；
5. 本设计将以模块的形式添加与树莓派日常工作系统之中。

设计内容：

1. 由于本设计主要用于采集数据，因此需对数据库的建立方式需要做一定的约定，而且可以自定义，以灵活应用于各种类型数据的挖掘；
2. Html标签属性和内容均可以实现读取并可指定读取内容范围，以及添加正则表达式匹配；
3. Html标签路径可实现自动匹配识别以简化操作流程；
4. 抓取方式暂定同TPaUrl v1.0之前版本相同，且多线程的起止位置可自定义设计；
5. 数据库的管理方式要灵活，暂定使用NoSql方式的Mongodb；
6. 为之后的数据处理留接口，可用于适配多种数据处理方式。