本周学习了UDF的使用，阅读了ccs2016的论文《Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence》

进展：上周从10000条文本里提取出的2400条URL对应到网站只有几十个，粗略看一下黑灰产网站非常少，所以说现有的舆情文本数据对于我来说没有太大用处。计划从相关论坛、贴吧等地方爬取链接，初步爬取了三万左右URL，质量还比较高，应该可以对应几百个黑灰产网站。下周尽快把待检测的网站名单确定。

困难：爬取刷单等网站时需要注册账号，绑定淘宝账号等。需要淘宝买家账号进行测试，不知道哪里申请。