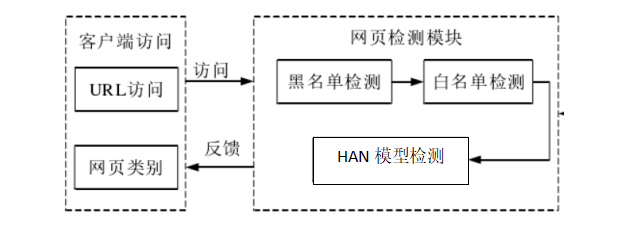
**需求实验：**

1. 数据集包括钓鱼数据集（20193条）和正常数据集（29782条），数据集url的特征有28维，以及url-ip和url-alink的对应关系。----已做
2. 构造包含url、ip、alink的异质图和url-ip-url和url-alink-url两个邻接矩阵（元路径）。
3. 用**HAN模型（pytorch框架）**对异质图url做节点分类，模型训练得到准确率、精确率、召回率、F1、ROC（出图）数据，同时包括两条元路径的权重，保存精度最好的模型。
4. 修改参数hidden layer（1，2，3，4，5，6，7）和embedding size（8，16，32，64，128，256）对ACC和F1的影响。
5. 将精度最好的模型制作钓鱼检测谷歌插件，流程主要有：首先是黑白名单过滤，有则直接反馈结果；无则进行模型的检测（这个模型检测的步骤有：将待检测的url的url地址、网页源码、ip用爬虫爬取下来并存储，数据处理包括找到url对应的alink和ip，输入到训练模型中做预测（预测分数大于0.5为正常（label=1），小于0.5为钓鱼（label=0））），最后将检测结果返回到页面。流程大致如下：



细节部分：

1. 黑白名单存储在数据库内，添加更新时间段，可更新数据库。
2. 系统做成Chrome 插件的形式，便于实时检测钓鱼网页
3. 系统的运行效果如下图，在访问疑是钓鱼网站的情况下会提示。（这个提示框会提示测的是哪一个地址，然后下面有一个检测按钮，检测结果在下面大字提示，如下图中“PHISHING”字样）
4. 有一个操作钓鱼检测（正常url 2个和异常url 2个）演示视频。

