

技术实训面试 10 问

杨嘉浩-求职意向数据分析实习生

1. 你的第一个实习项目是怎样的？

答：首先，我们需要解决的是朋友圈宣传运营图片审查的问题，对其我们做了三个方面的工作，分别是制定标准、设计检验算法以及 web 平台搭建。首先通过研究各个图片的规律得到出了六条审查标准，其中为了更好地发现不同一图片同一位置的部分，我们采用了 labelhub 进行了数据标注；其次根据标准实现了自定义的审查算法，其中涉及到图片文本检测所以我们调用百度云 ocr 接口。第三基于我们希望能够让工作人员在网站平台上审查图片，因此搭建了一个 web UI 平台，其中采用了 builder 组建框架，以及 tornado 技术实现。

2. labelhub 对你们项目的作用是什么？

答：首先 labelhub 是一个数据标注的平台，由于我们在制定审核标准的时候需要对每张照片进行一个细分部分的分析。而 labelhub 能够帮助我们在不同的图片上切割不同的部分，因此让我们便于总结其中的规律，更好地得出了图片审查规则。

3. 你们得到的标准制定有哪些？

答：总共有六点。第一点是检测该图片是否为朋友圈的详情页，因为需要得到详细的审核的要素。第二是审查图片是否为正常的本人的全页面。第三是该朋友圈发送时间是否是在晚上 7:00 之后，第四是留存的时长必须为 4 小时以上，第

五是朋友圈正文以及评论话术要符合要求。第六是分享的链接必须为有效链接。

4. 这些标准的依据是什么？

答：两个方面，一方面通过我们业务的需求，另一方面是对获取图片的分析，比如说审查时，有些非本人的图片会出现两个通知栏以及没有删除的字样，我们希望把这两点当作是审核的标准，又比如说制定发送时间为晚上 7:00 之后，是因为这段时间是朋友圈流量高峰期，我们希望在这时间点让更多的人看到。

5. 你们有对图片进行预处理吗？

答：我们通过两个方面对图片进行了预处理。第一是我们获取到了图片的信息，比如说长宽高尺寸属性类别等基本信息。第二是我们对图片进行了查重，以保证图片质量的正常。

6. 你们的自愿役的检测算法大概是怎么样的呢？

答：在检测算法上我们分三大部分进行了审核。第一是检测图片是否为本人，这里我们通过了查找是否为一条通知栏，以及是否有删除字样来进行鉴别。第二次检测内容是否与要求符合，其中需要比较正文、链接、评论内容部分。第三是对时间进行一个鉴别，我们会鉴别他的转发时间是否为 7:00 后，以及与图片通知栏时间进行对比，得到了验证结果。

7. 在 web 平台搭建前有产品人员参与吗？其中你们在产品的部分是怎么做的呢？

答：是这样的，因为我们部门有比较擅长产品的同事，由他们在项目产品部分上进行了设计。产品部分大概是首先设计了功能的范围，其次是对功能之间的交互上次进行了研究与设计，最后通过 builder ui 组件平台，设计出了高保真的原型图，在这个过程中还设计了数据库的 e-r 图并产出了一份功能设计技术报告文档，为平台的实现搭建作下来基础。

8. 你们是怎么实现这个 UI 平台搭建？

答：首先基于功能分析技术报告，对于技术的落地点有深入的明确。接下来通过 tornado 技术建立了服务器，把我们的功能变成一个真正可供用户使用的服务端。其中实现了许多 builder 平台交互的功能设计。

9. tornado 对你们项目起到了哪些帮助？

答：首先 tornado 是一个连接客户端的服务器，他能把接收到的 get post 请求返回呈现给客户端。因此，在这个项目里，我们借用 tornado 实现了用户登录、项目创建、图片上传、图片审查等功能，通过 web 实现了对用户的实时反馈。

10. 这个项目的结果是什么？你们给公司带来了多少的利益。

答：项目的结果是从零到一地完成了朋友圈宣传图片审查系统，搭建出了一套能供用户操作使用的 web 平台。提高了公司运营业务的审核效率。