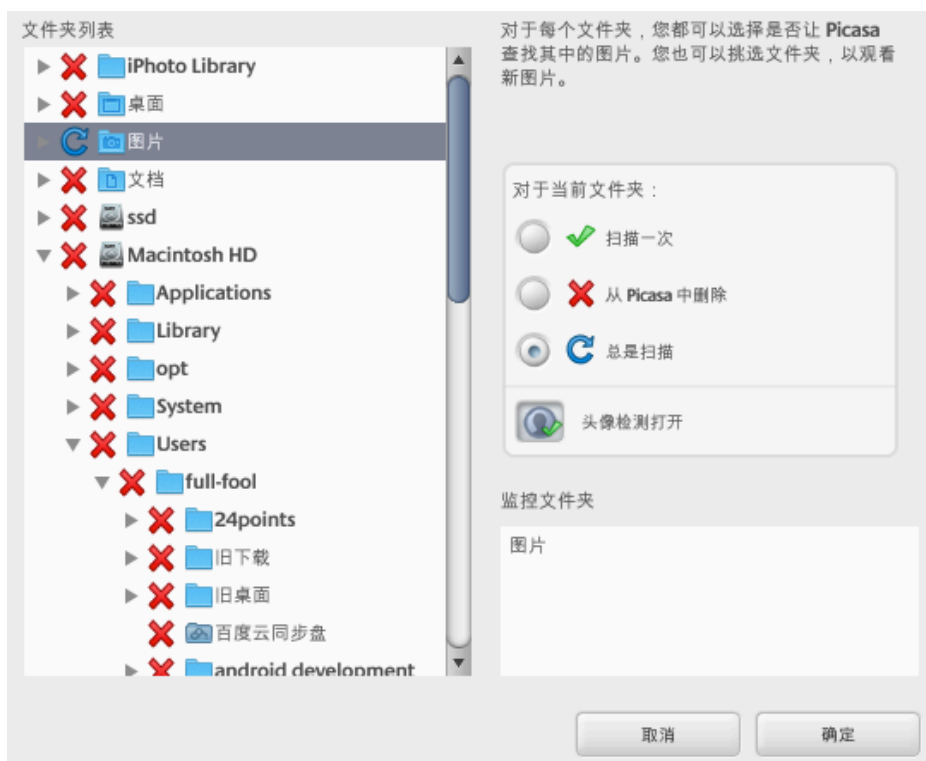


Grouping 流程

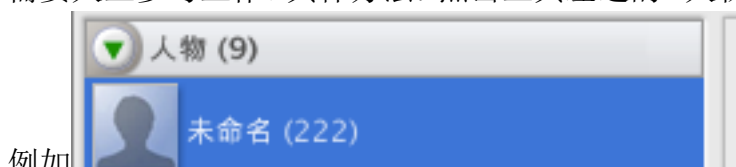
作者：崔逸卿

1. 人工初筛。人工筛选掉一些使用价值不高的图。例如：
 1. 没有人脸的照片。
 2. 没有主题的照片。为拍摄者随手拍，画面中没有主要人物，或者一看就是拍了一堆路人，不包含主要人物的图片。
 3. 画面主题不是人的图片。例如图片主要拍摄花花草草，却在画面的角落里乱入了几个人脸。
 4. 无意义的人脸。比如拍摄的电视机、电影、海报、广告中出新的人脸。
 5. 质量不高的人脸。比如图片中的人物太模糊、90° 侧脸、脸被遮挡的面积过大（例如超过 40%）等。
2. 将图片导入 **picasa**。简单起见，我们在桌面新建一个文件夹，将完成步骤 2 之后的图片文件夹全部放进去。然后打开 **picasa**，设置扫描图片路径。具体：工具 -> 文件夹管理器，打开画面如下：



为了仅导入我们需要的图片，此处将所有的其他文件夹全部选择红色的叉(从 **picasa** 中删除)，仅仅对之前新建的图片文件夹选择“总是扫描”，然后确定。此时 **picasa** 会开始扫描图片并检测人脸（如果左部浏览栏中有无关文件夹，可以手动选择，右键“从 **picasa** 中删除”）。

3. 人工聚类。虽然工具会协助我们聚类一部分人脸，但其中有一定比例的错误，需要人工参与工作。具体方法：点击工具左边的“人物”标签下的“未命名”，



例如，然后在右边选择“展

开群组”，例如



（如果没有展开群组，只有“按头像分组”和“显示忽略的头像”，则说明已经展开了），此时会发现所有的人脸已经按照一定的顺序归类在一起了。此时随便选择一张图片，为他命名并创建联系人，则所有的被软件识别为同一个人的人脸就自动归到了这个名字下。有许多软件不能确认的图片需要人手工归类。注意：不同的人要用不同的名字，名字没有固定的规则，尽量使用字母和数字，不能重复。对于那些出现频率很少的“路人”，可以选择忽略他，不为他归类。最终我们需要的结果是每一个主要人物的头像都被聚类到了一起，并且有一个独一无二的名字。

4. 根据数据库生成 xml。进入到工具 `exportpicasa-0.4.1-win32` 目录下，运行工具 `exportpicasa.exe`。直接运行 `exportpicasa.exe` 可以得到工具的使用方式。以我的电脑 win7 为例，运行的命令为 `exportpicasa.exe -d "C:\Users\username\AppData\Local\Google\Picasa2" -o "C:\Users\username\Desktop\exportpicasa-0.4.1-win32"`，其中 -d 后的参数为数据库文件所在位置（AppData 可能为隐藏目录，因此可能需要设置查看所有文件夹），-o 后的参数表示生成的 `index.xml` 文件存放的目录。
5. 根据 `index.xml` 生成结果表格。将生成的 `index.xml` 移动到和脚本 `processXml.py` 同样的目录下，然后运行 `processXml.py` 脚本。程序会要求输入结果文件前缀，输入源文件夹名字即可。比如这个 `index.xml` 是根据图片文件夹 `renren_123456` 中的图片聚类生成的，则输入 `renren_123456`。程序会输出对应的 `csv` 文件。这就是我们最终需要的文件。

建议：建议每次处理一个文件夹，即在步骤 2 中每次只导入一个文件夹，并且导入之后将之前的结果删除，将之前处理的文件夹“从 `picasa` 中删除”，保证每次处理的图片只来自于同一个文件夹。