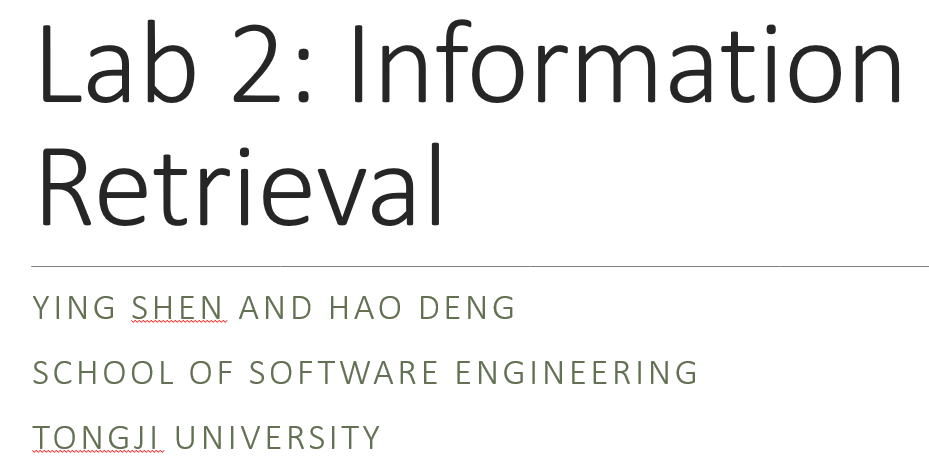
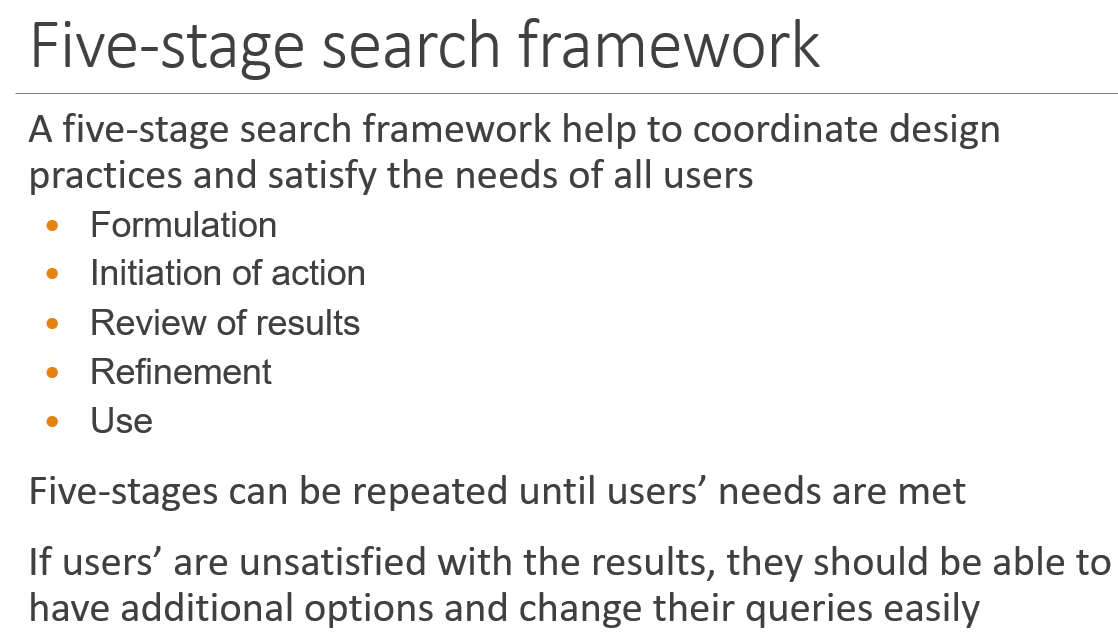
**P1：**



这是我们的第二次个人项目

**P2：**



本次个人项目侧重对五阶段搜索框架的练习

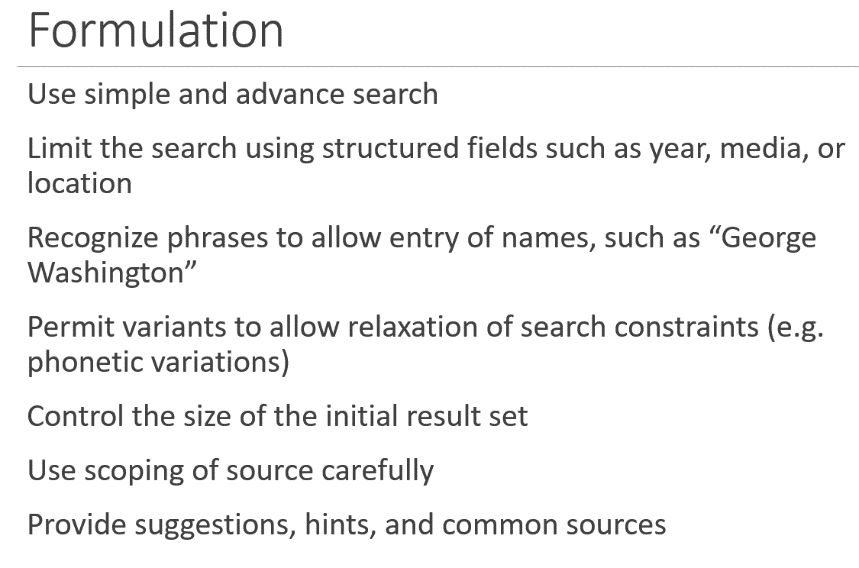
我们再回顾一下：搜索主要可以分为五个阶段

1. 构想。表达搜索
2. 发起动作。启动搜索
3. 评审结果。阅读结果
4. 细化。构想下一步
5. 使用。编辑或传播结果

信息搜索本身是一个迭代过程。因此，设计信息搜索系统，也可以按照这五个阶段，在设计时重复多次，直到设计的系统满足用户需求。

用户不一定总能在信息搜索时，看到所有的五个阶段。但是如果用户对结果不满意，要让用户可以获得额外的方法来改变他们的搜索查询。

**P3：构想**



构想阶段：

用户会构想：是需要使用简单搜索还是高级搜索。那么设计的时候要根据用户需求，考虑提供那种功能，或者两种功能都兼顾

也要考虑到，为了方便用户具体化他的搜索需求，是否需要提供结构化的字段，来帮助用户对自己的搜索进行限制

如果用户输入短语作为关键词的时候，可以考虑提供智能化补全或短语识别功能，比如能够智能地分辨搜索“乔治·华盛顿”还是华盛顿这个地名

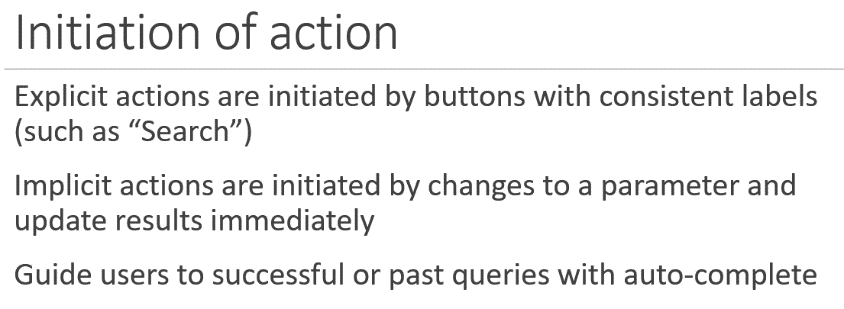
构想阶段，也要允许放宽搜索的约束条件，考虑到用户错别字、近义词、地方口音等的影响，能够在搜索结果中响应这种变化

同时，又要能帮助用户在这个阶段控制住初始的搜索结果数量

也可以为用户提供信息来源范围的细化选择功能，比如只搜索近3年，或者只搜索专利数据库

要能为用户提供构想搜索的建议、提示，以及保证常见的信息来源

**P4：发起动作**



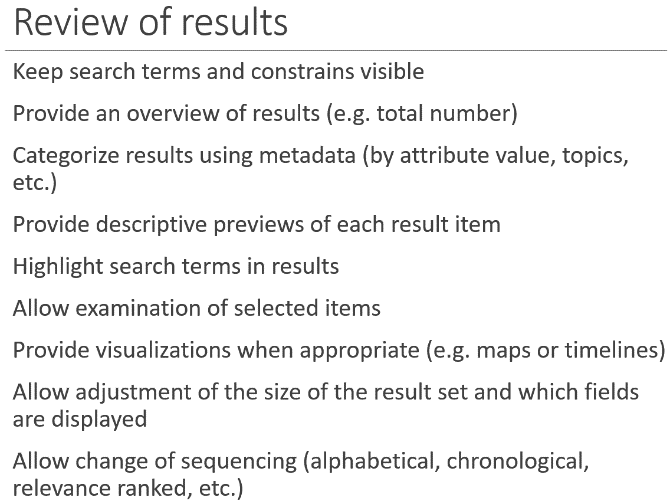
发起动作阶段：

典型的，要给用户设计搜索按钮来发起搜索动作。搜索按钮要跟大部分的搜索引擎具有一致性的设计

发起动作也可以通过改变搜索参数和立即更新结构等功能，来触发隐式搜索动作

发起动作也可以用自动完成功能等，来帮助用户重复过去的成功搜索

**P5：评审结果**



评审结果也就是查看结果阶段：

在结果呈现的界面，也要保持搜索词和约束条件可见。这样可以帮助用户迭代搜索时，提高效率

呈现出来的结果要提供概要，比如搜索结果总共多少条

搜索出来的结果可以通过元数据进行分类，比如通过属性值、主题等

每个搜索出来的结果，能够显示图像或文字描述性的预览

搜索出来的结果中要能为用户凸显出搜索的关键词

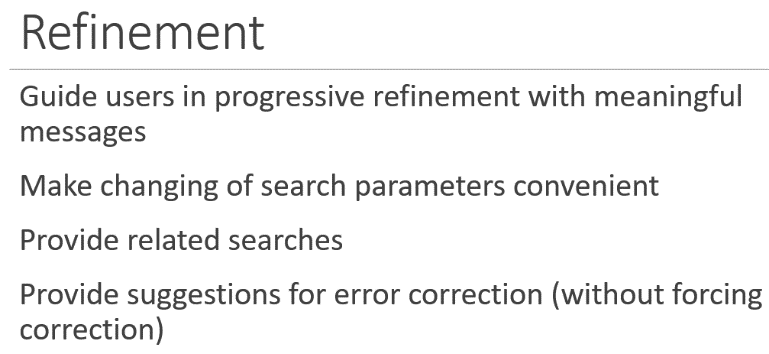
要能让用户能够查看选定的条目

适当的情况下，可以给搜索结果提供地理位置信息或时间信息

要能允许用户调整结果数量的大小，或者范围字段

要能允许用户更改结果的排序（比如按字母顺序排序，按价格高低排序，按时间先排序，按相关性排序等）

**P6：细化**



细化阶段：

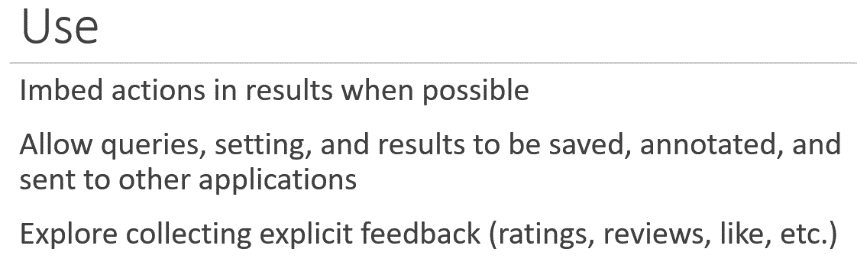
应该引导用户逐步完善有意义的检索信息

应该提供搜索参数更改功能，让用户可以方便地更改搜索参数

除了最直接相关的搜索外，可为用户提供相关的搜索

能够发现用户的错误搜索，并为错误提供修正建议，但是不要强迫用户修正

**P7：使用**



搜索结果使用阶段  
  
可以考虑将搜索操作嵌入到结果中，方便用户对搜索结果跟踪或掌握来源

可以允许用户保存查询的关键词，设置的搜索范围和获得的搜索结果，并提供说明，以及能够把这些信息发送到其他应用保存或使用

使用的阶段可以为用户提供显示反馈渠道（比如评级、评论、喜好等），这些使用情况可以为搜索系统的隐式搜索或改进设计提供反馈

**P8：我们来看一下百度网页搜索的例子**

**构想方面**，它的简单搜索功能非常易用

在搜索框中也可以智能地识别关键词短语

**可以很方便地选择信息来源范围**，也可以在**设置那里设置特殊的搜索字段**

**发起搜索动作方面，**它提供了一个显著的搜索按钮，也可以**在设置那里通过改变参数来触发隐式的搜索动作**

**P9：**

**评审结果方面，**我们可以看到它保证了搜索词可见，

提供了结果的概览，

允许对结果提供特定主题，

搜索结果中也突出显示了搜索关键词

给出了搜索结果的来源

给出了搜索结果的描述性预览

**P10：**

**细化结果方面，**

我们可以在百度的搜索工具上，改变搜索参数

**P11：我们来看一下百度图片搜索的例子**

**构想方面，**用户既可以输入关键词，又可以输入图片

也提供了关键词自动补全功能

提供了最热的搜索推荐

发起搜索动作的按钮与网页搜索类似

**P12：如果用用户界面作为关键词搜索，可以看到搜索结果**

也在细化方面提供了**相关搜索**功能

**P13：细化方面**

还提供修改搜索参数的功能，图片筛选

**P14：我们再看一下文档搜索的例子**

这是我们学校图书馆的图书检索系统

除了简单搜索外，图书检索系统的**高级搜索**会比网页搜索引擎更显著

**P15：评审结果方面**

除了提供结果概览，也提供了基于元数据的结果分类检索。这是典型的分面搜索设计

图书检索系统的结果排序功能通常也会设计地比较显著

也通常会提供关键词突出显示的功能

**P16：细化结果方面**

支持用户**在结果中进行搜索**，让结果更符合需要，也支持重新搜索来更改搜索需求

**P17：搜索结果使用方面**

图书或参考文献的搜索都有显著的收藏结果、提供评价、提供搜索信息的功能。这些功能有利于用户对搜索结果跟踪或掌握来源

**P18：如果用Python实现信息检索**

可以安装Python的轻量级Web应用框架Flask。Flask本身相当于一个内核，其他几乎所有的功能都要用第三方的扩展来实现。

flask-HTTPAuth是一个简单的扩展，它简化了使用Flask路径的HTTP认证使用。

NumPy是Python的一个用于科学计算的基础包。它提供了多维数组对象，多种衍生的对象（例如隐藏数组和矩阵）和一个用于数组快速运算的混合的程序，包括数学，逻辑，排序，选择，I/O，离散傅立叶变换，基础线性代数，基础统计操作，随机模拟等等。

TensorFlow 是世界上最受欢迎的开源机器学习框架，它具有快速、灵活并适合产品级大规模应用等特点。

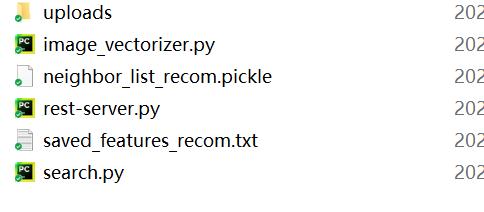
SciPy是一个开源的Python算法库和数学工具包。 SciPy包含的模块有最优化、线性代数、积分、插值、特殊函数、快速傅里叶变换、信号处理和图像处理、常微分方程求解和其他科学与工程中常用的计算。

Imageio是一个Python库，提供了一个简单的接口来读取和写入各种图像数据，包括动画图像，视频，体积数据和科学格式。

matplotlib是Python的绘图库。它提供了一个面向对象的API，用于使用通用GUI工具包将绘图嵌入到应用程序中。

sklearn 是一款开源的Python 机器学习库，它基于NumPy 和SciPy，提供了大量用于数据挖掘和分析的工具，以及支持多种算法的一系列接口。

**P19：**



**先运行image\_vectorizer.py，**它通过谷歌的inception-v3卷积神经网络模型处理各个图像数据，完成图像分析和对象检测等任务，并把卷积神经网络中的瓶颈层向量收集和存储起来。

瓶颈层：https://zhuanlan.zhihu.com/p/98692254

注意，在image\_vectorizer.py中，要相应地修改数据集的路径

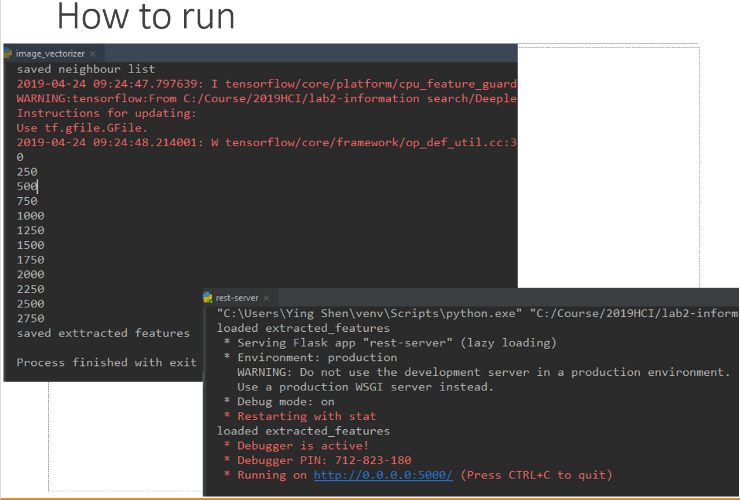
运行后，会产生两个文件，即image\_list.pickle和save\_features.txt。把它们放在存有search.py脚本的文件夹里。

运行rest-server.py，启动服务器，为本项目拟设计的用户界面提供基于flask的REST服务

一旦服务器启动了，通过访问网址（例如）0.0.0.1:5000就获得用户界面。

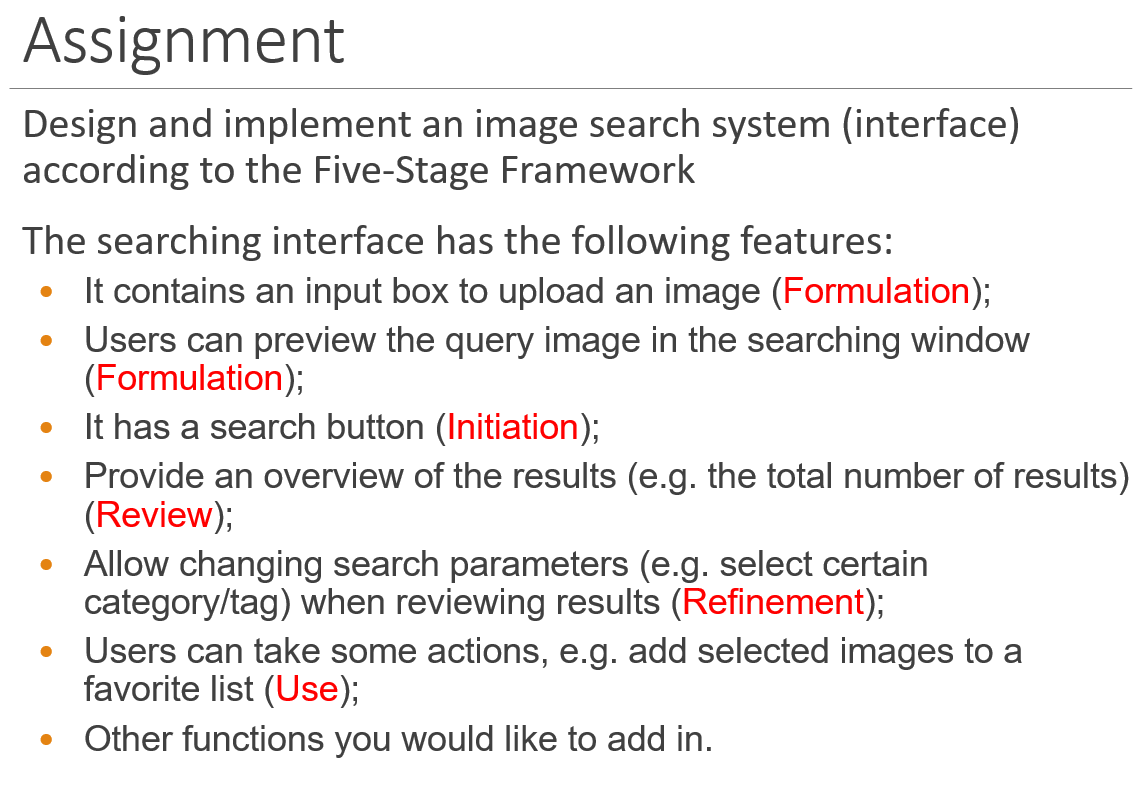
现在如果你上传任何文件，可以看到9张类似图片。你可以将K值从9改为任何值，但不要忘记相应地更新html文件

**P20：**



Xxxx

**P21：项目任务**



采用五阶段框架，设计一个图片搜索系统界面

**搜索界面要求具有以下特性：**

**Xxx**

**P22：需要用英文提交1个报告**

