大作业提供三道选题，请从选题中选取任意一题，实现指定技术要求。

最终成绩由基础实验完成分值（80%）+自主设计额外亮点（20%）构成。

一、 学生信息检索系统

1. 题目说明

本题提供一份数据集（oulearning.zip），包含了某线上 MOOC 平台若干年的学生选课、作业成绩记录以及课程资源目录等各维度信息，详细文件及字段说明见数据集内“data_info.md”文件。

请实现一个界面化信息检索系统，读入数据并支持用户按照各种检索条件展示符合条件的信息。

2. 作业要求

2.1.数据处理与读取

从全部数据文件中读取完整数据，并对数据中可能存在的缺失值等异常进行处理。

2.2.数据结构建模

本数据集涉及多维度信息，请根据实际需求设计合适的数据结构，为对象或者辅助类进行建模。设计哪些数据结构，分别在何时使用自行按需求确定。

2.3.实现数据的条件检索

用户可以从界面中检索指定的学生、课程或作业信息；检索某一课程全部学生；检索某一学生最终平时成绩；检索某一学生的课程资源历史交互记录等等。请设计多种检索条件，并实现按照检索条件准确返回符合条件的记录。

2.4.界面展示

提供一个可交互界面，支持用户实现上述各检索功能，并展示检索结果。

3. 技术要点与提示

1. 本数据集原始数据量较大，可考虑如何精简**存储空间**，例如使用“枚举”来减少字符串等类型字段的存储空间等。
2. 定义**合适的数据结构**来存储或检索数据。而且是否有必要在程序运行全程使用 class 表示每一份数据？*关于 object __dict__的额外内存消耗：https://zhuanlan.zhihu.com/p/88589720*
3. 如何实现一个高效且准确的**检索算法，例如使用 B+树管理信息**。
4. 界面简洁可用即可，美观度不做要求。

二、 爬虫检索系统

1. 题目说明

爬虫是一种很常用的获取数据的方法，请**任意选取**自己感兴趣的信息平台网站（如豆瓣读书、电影等，或微博话题等，类别不限），使用 python 相关库函数，从选取的网页爬取并保存对象的信息（如电影名称、简介、评论等），以及页面上的相关链接（如相关同类型电影、出演影星等）。

将爬取信息保存到本地后，请实现一个检索展示系统，支持用户从数据中检索对应的信息并展示。

2. 作业要求

2.1.数据爬虫

学习爬虫相关技术，爬取自选网站的相关信息并保存。**数据量**需要足够支撑现场演示的检索需求。完成数据爬取后，请将数据保存为本地文件供后续步骤使用。

2.2.B+树检索

由于爬虫保存的本地文件通常数量较多，全部一次性读入内存并不合适。因此需要按需读取数据资源，在程序启动时只需读入资源文件路径、数据标识信息（例如电影名称或 ID）等关键信息即可。

基于 B+树设计一个检索算法，根据指定的搜索条件（例如 ID）搜索对应资源保存路径，读取对应资源。

2.3.数据结构建模

请设计合适的数据结构来描述检索项，对爬取的信息进行建模。

2.4.界面展示

提供一个可交互界面，支持上述检索功能，并展示检索结果。

3. 技术要点与提示

1. 爬虫数据量足够支撑现场演示的检索需求即可，如果爬取了图片音频等多媒体资源，可适当减少数据量以节省本地空间占用。
2. 更多爬虫技术问题可见文档“爬虫 FAQ”。
3. 关于**第三方库**的使用：由于爬虫是一个实际的工程技术，因此鼓励同学使用成熟的开源库来解决爬虫中遇到的问题，如果能够应用合适的开源库工具解决开发中的问题，例如使用 selenium 来模拟浏览器等，亦可作为大作业的闪光点。

三、复现基于蒙特卡洛算法的 Ising 模型

1. 题目说明

本题目提供一份英文文献及代码附录（ising.pdf）。

Ising 模型是物理学中一种模拟物体状态变化的常用模型。最初 Ising 模型用来模拟金属磁场的随环境变化过程。本实验需要基于 python 语言，**通过蒙特卡洛的随机化算法进行 Ising 模型的迭代模拟**。附件 pdf 包含了 Ising 模型的推导过程以及 matlab 实现。

在二维 Ising 模型中，假设有 N*N 的粒子均匀以矩阵形式摆放，每个粒子存在一种磁性，用+1 或-1 表示。粒子的磁性会与周围紧邻的另外四个粒子的磁性向交互，当两个粒子彼此相邻，且彼此磁性不同时，则处于不稳定的高能态，反之则处于稳定的低能态。此外，系统环境可能会存在一个整体的磁场，与之相反的粒子也会有较高的能量。

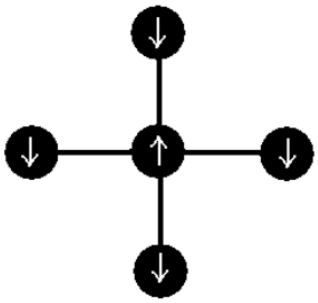
综上所述，粒子的能量由两部分决定：相邻其他粒子磁场以及系统总体磁场。从此可推导出系统中全部 N*N 个粒子的总能量，表示为：

$$H = -J\sum_{\langle ij\rangle} S_i S_j - B\sum_i S_i$$

其中第一项为每个粒子与相邻粒子之间磁场差异带来的能量，$< i,j >$是全部相邻的粒子有序对，$\sum S_i S_j$为对相邻粒子磁性乘积求和。J 为系数，$J > 0$时表示粒子间差异越大，系统能量越高，即越不稳定。可见若全部粒子磁性均相同，则此项能量会达到最小。

第二项为每个粒子与总体磁场之间差异带来的能量，B为系数。

为了使系统总能量降低，达到稳定状态，粒子会随机地反转自己的磁场，以达到更低的能量。为方便起见，可假设$B = 0$，只考虑相邻粒子对自身的影响。如下图，假设$J = 1$，中心粒子的磁场为+1，其余均为-1，此时该粒子的能量为$-(-1-1-1-1) = 4$，而反转磁场后，能量变为-4，因此更加稳定。



基于蒙特卡洛的 Ising 模型在实验中首先随机为每个粒子赋予+1 或-1 的磁场方向，随后在实验过程中会进行多次迭代（原实验设置每次实验迭代次数为10^3），每次迭代过程：

1. 计算每个粒子的能量。
2. 基于蒙特卡洛算法，随机选取部分粒子反转其磁场。

注意在蒙特卡洛算法中，不同能量状态的粒子反转的概率不同，能量越高，反转概率越高。从而实现迭代过程中能量的下降，系统逐渐达到稳定状态。

2. 作业要求

实验复现

请阅读附件文献，结合上述内容以及文献附录代码，**基于 python 复现实验**。其中“ising.m”为每次实验的迭代过程。

多次进行实验并记录实验结果，利用 matplot 绘制附件文档中 Figure 2-3 的实验结果展示图例。其中 Fig. 2 表示不同参数 J 下，模型经过10^3次趋于稳定后的粒子平均能量的散点图；Fig. 3 表示粒子平均磁场大小的散点图。

本作业考察创新能力，开发过程中不组织本项作业的答疑，同时鼓励同学自主阅读原文文献，理解原推导过程与公式，在实验中提出自己想法，例如针对实验结果绘制其他图表，尝试在 Ising 模型中引入更多参数等。