

编程作业 2: 网络编程

- [30%] 利用文件随机访问方法，设计一个如下文件内容：通过随机数生成器（设定种子值为244），生成2个不同的0.0-1.0之间double数值对<d1, d2>。重复生成该数值对<d1, d2>，终止条件为d1和d2均大于0.91。将上述生成的数值对<d1, d2>按照如下协议保存到文件（名字为h2q1.dat）中：
 - 将所有随机生成的数值对<d1, d2>，按照其中的数值大小分组：若d1和d2均小于0.46，则保存到第一个分组区域；若d1和d2均大于0.72则保存到第三个分组区域；否则其他数值对则保存至第二个分组区域。所有分组内的数值对，则按照数值对的随机生成顺序保存。
 - 将所有double数值按照字节数组byte[]，并按照字节数组形式保存到文件；
 - h2q1.dat文件的格式（见下图）：
 - 头部：数值对个数...
 - Group1所含数据，Group2所含数据，Group3所含数据



根据上述要求，输出如下任务：

- 统计数对随机生成和文件保存的整体时长，时长越短则效果更优、评分越高；
- 统计数对中大于0.5的double数值个数（数对中无论d1或者d2大于0.5均计入统计）和数对中大于0.5的double数对个数（数对中d1和d2都大于0.5才计入统计），时长越短则效果更优、评分越高；

先保存在磁盘，再random；不是在内存

可以牺牲空间，不用等所有grp都生成数据

- [30%] 通过编制http协议客户端程序访问4个种子网站主页：www.tongji.edu.cn，www.pku.cn、www.sina.com.cn和www.mit.edu，解析html页面，从中选择至多6个外部url超链接（要求该外部url域名不同于当前机构，即要求域名url最右后缀至多有1个或2个相同的子域名，例如同济图书馆www.lib.tongji.edu.cn与www.tongji.edu.cn共享三个相同的最右后缀子域名cn、tongji和edu，不合要求、不能作为外部地址url）。在到达该外部url地址后重复上述操作，访问最多4跳范围的外部地址url（要求不包含起始种子地址）。在上述访问外部url过程中，若外部url为重复地址，则无需从该外部url解析html页面和选择其中的url。

根据上述种子网站url和外部url，通过一个可视化工具代码绘制一个有向图，其中图节点为所有访问的url，节点之间的有向边为超链接关系，并打印(1)统计该有向图的节点和有向边数量、包含in-degree（入边）数量最多的节点url；(2)上述访问url的总体耗时（单位：秒）

3. [40%]完成一个client和多server形式的文件上传和下载的socket网络程序：Client在指定需要传输文件（比如100MB），分块该文件（比如1024KB）讲分块文件随机传输至多个servers（比如K=3），服务器程序在接受请求后将数据块保存至本地文件系统。为了确保客户端可以下载已经传输的文件，要求在服务器端记录一个文件名+文件块与所在服务器位置的映射表，以确保在完整文件的上传之后，客户端可以通过查阅该映射表请求对应的服务器来下载文件。此外，在client与server之间的网络连接发生故障时，要求文件传输具备续传功能，不必文件起始点重新传输，以优化网络带宽资源。

根据上述功能描述，要求客户端程序通过标准输入console的形式接受如下指令：

- 传输(put)和下载(get)指令：通过指定上传文件名称和下载文件保存位置，在完成传输和下载指令后，并输出所需时间；
- 具有续传传输(cput)和下载(cget)指令：上述put/get指令不具备续传功能，cput/cget则可以在某个网络通道断开后续连接后具有断点续传能力；在完成传输和下载指令后，并输出所需时间（该时间不记录网络断开时长）；
- 查询(check)指令：检查K个服务器的状态，包括网络连接是否通达、指定文件的存储映射信息（该文件数据块的所在服务器地址、所在服务器所包含的数据块个数、数据块小从至大的ID序列）。

提交要求：

- 1) Deadline：2023-11-19 23:59
- 2) 每题包括3个文件：源代码文件目录（学号-hw2-q1/2-src），执行/运行文件目录（学号-hw1-q1/2-run）以及执行截屏文件(学号-hw1-q1/2-screen)，其中执行截屏文件要求截屏记录console输入脚本文件并输出结果。
- 3) 在 canvas 上提交最终压缩文件包（学号-hw2.zip）