

2021-2022 秋 期末考试

'''

Author: XiangxinZhong

Date: 2021年12月28日18:36:43

LastEditors: XiangxinZhong

LastEditTime: 2021年12月28日18:36:48

Description: 2021秋季期末考试总结

'''

一、操作系统

1. 题型

仅包含能够回忆起的题目

选择题

1. openEuler;
2. OS: 何老师重点

填空题

简答题

1. 线程与进程的区别。
2. 微内核内核结构定义、优势、应用的OS? (最后一个我没写出来)
3. 简述死锁的4个必要条件。
4. 如何看待openEuler的新特性, 简述4点。

综合题

1. 信号量的应用: 前驱图, 课本P113;
2. 分页存储管理: TLB时间计算, 课本191例题;
3. 磁盘调度算法: FCFS, SSTF, SCAN移动磁道计算;
4. 进程周转时间: 计算非抢占式和抢占式SJF的周转时间;
5. 综合题: TLB的综合使用, 据说是考研真题;

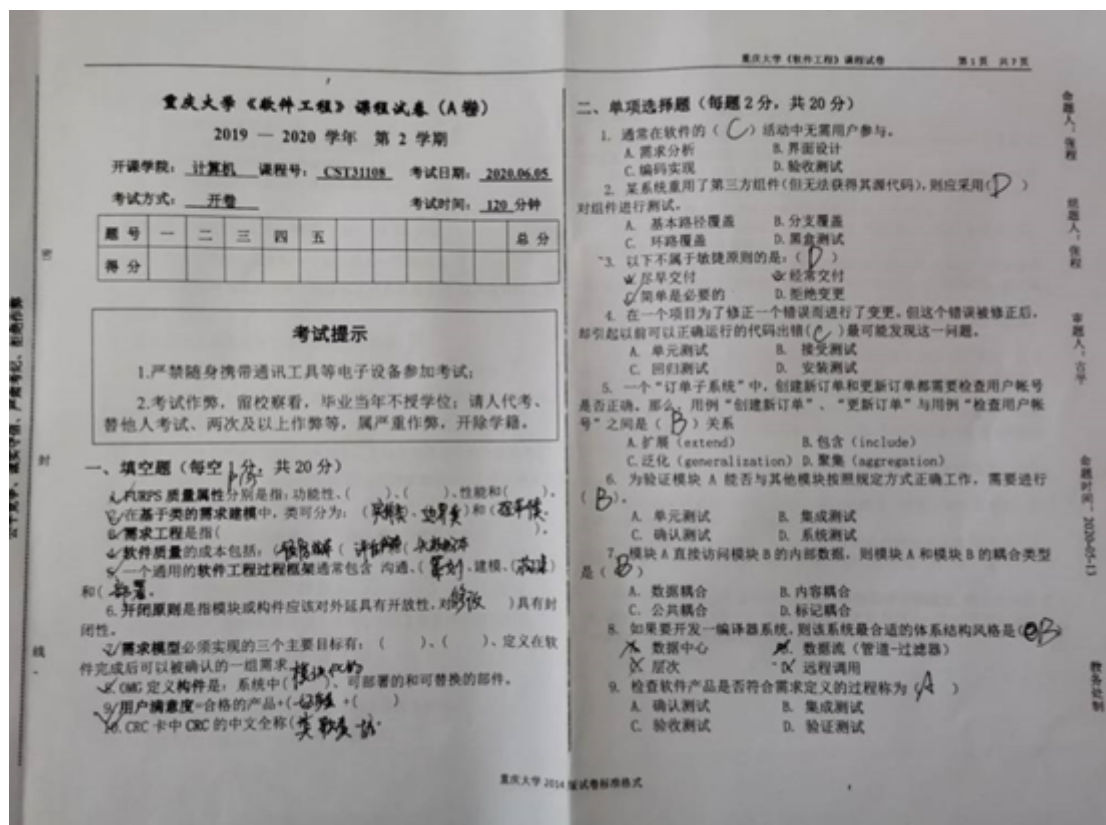
2. 复习建议

1. 课本+PPT: 根据何老师画的重点复习;
2. 有时间可以做点计算题

二、软件工程

1. 题型

参考2020年期末



10. 数据流图的主要作用在于 (D)

- A. 描述数据对象及其关系。
B. 描述某加工内部的处理逻辑和算法。
C. 描述系统如何对外部事件响应处理的。
D. 描述数据如何在系统中变换和流动的。

三、判断题 (标记T或F, 每题1分, 共10分)

1. DFD中的每个加工至少有一个输入流和一个输出流。T
2. 内聚性和耦合性是度量软件模块独立性的主要准则, 软件设计时应力求“低内聚, 高耦合”。F
3. 软件测试的目的是为了证明程序没有错误。F
4. 接口分离原则强调多个客户专用接口比一个通用接口要好。T
5. 状态图展现了多个(不同)类的对象彼此间的协作和状态变迁过程。T
6. 确认是指确保软件正确地实现某一特定功能的一系列的活动。F
7. 软件可用性是指某个给定时间点程序能够按照需求执行的概率。T
8. 单元测试中不需要桩和驱动模块, 因为此时被测模块间是彼此独立的。F
9. 软件就是计算机程序。F
10. 不成熟的软件过程往往难以开发出高质量的软件产品。T

四、简答题 (每题4分, 共20分)

1. 请简述 David Hooker 提出的 7 个关注软件工程整体实践的原则。
2. 请简述需求工程包括的 7 项明确任务。
3. 请简述信息隐藏原则中的“隐藏”的含义。
4. 请简述 Theo Mandel 关于界面设计的 3 条黄金规则。
5. 请简述软件质量保证 (SQA) 包括的内容。

五、应用分析题 (每题10分, 共30分)

1. 网上药店允许顾客凭医生开具的处方, 通过网络在该药店购买处方上的药品。该网上药店的基本功能描述如下:

- (1) 注册: 顾客在买药之前, 必须先在网上海店注册。注册过程中需填写顾客资料以及付款方式(信用卡或者支付宝账户)。此外顾客必须与药店签订一份授权协议书, 授权药店可以向其医生确认处方的真伪。
- (2) 登录: 已经注册的顾客可以登录到网上海店购买药品。如果是没有注册的顾客, 系统将拒绝其登录。
- (3) 录入及提交处方: 登录成功后, 顾客按照“处方录入界面”显示的信息, 填写开具处方的医生的信息以及处方上的药品信息。填写完成后, 提交该处方。
- (4) 验证处方: 对于已经提交的处方(系统将其状态设置为“处方已提交”), 其验证过程为:
 - ① 核实医生信息: 如果医生信息不正确, 该处方的状态被设置为“医生信息无效”, 并取消这个处方的购买请求; 如果医生信息是正确的, 系统给该医生发送处方确认请求, 并将处方状态修改为“审核中”。
 - ② 如果医生回复处方无效, 系统取消该处方, 并将处方状态设置为“无效处方”。如果医生没有在 7 天内给出确认答复, 系统也会取消该处方, 并将处方状态设置为“无法审核”。
 - ③ 如果医生在 7 天内给出了确认答复, 该处方的状态被修改为“准许付款”。
- (5) 对于通过验证的处方, 系统自动计算药品的价格并邮寄药品给已经付款的顾客。该网上药店采用面向对象方法开发, 使用 UML 进行建模。

下图给出了“处方”的部分状态图。根据说明中的描述, 用上文中的文字, 补出图中缺少的 S1~S4 所对应的状态名以及 (1)~(3) 处所对应的迁移 (transition) 名。请在答题纸上直接作答, 不要重新画图。



2. 某高校欲开发一个成绩管理系统, 记录并管理所有选修课程的学生的平时成绩和考试成绩。其主要功能描述如下:

- (1) 每门课程都有 3 到 6 个单元构成, 每个单元结束后会进行一次测试, 其成绩作为这门课程的平时成绩。课程结束后进行期末考试, 其成绩作为这门课程的考试成绩。
- (2) 学生的平时成绩和考试成绩均由每门课程的主讲教师上成绩管理系统。
- (3) 在记录学生成绩之前, 系统需要验证这些成绩是否有效。首先, 根据学生信息文件来确认该学生是否选修这门课程, 若没有, 那么这些成绩是无效的; 如果他的确选修了这门课程, 再根据课程信息文件和课程单元信息文件来验证平时成绩是否与这门课程所包含的单元相对应, 如果是, 那么这些成绩是有效的, 否则无效。
- (4) 对于有效成绩, 系统将其保存在课程成绩文件中。对于无效成绩, 系统会单独将其保存在无效成绩文件中, 并将详细情况提交给教务处。在教务处没有给出具体处理意见之前, 系统不会处理这些成绩。
- (5) 若一门课程的所有有效的平时成绩和考试成绩都已经被系统记录, 系统会发送课程完成通知给教务处, 告知该门课程的成绩已经齐全。教务处根据需要, 请求系统生成相应的成绩列表, 用来提交考试委员会审查。

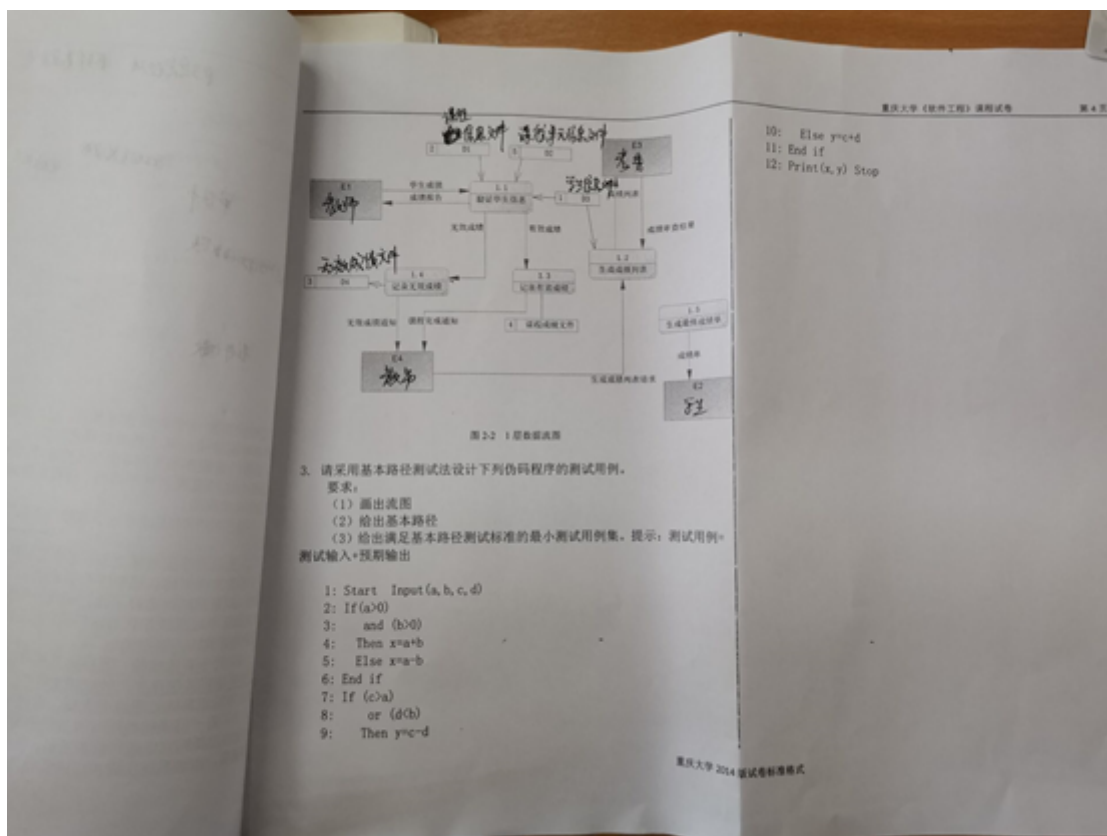
- (6) 在生成成绩列表之前, 系统会生成一份成绩报告给主讲教师, 以便核对是否存在错误。主讲教师须将核对后的成绩报告返回系统。
- (7) 根据主讲教师核对后的成绩报告, 系统生成相应的成绩列表, 提交考试委员会进行审查。考试委员会在审查之前, 上一步的成绩报告已经系统, 对于所有通过审查的成绩, 给出学生信息文件, 系统将会生成最终的成绩单, 并通知每个选课学生。

现采用结构化方法对这个系统进行分析与设计, 得到如下图 2-1 所示的顶层数据流图和图 2-2 所示的 1 层数据流图。

请根据说明中的描述, 用上文中的文字, 补出图中外部实体 E1~E4 的名称, 数据流 D1~D4 的名称。另外, 图 2-2 中缺失了两个数据流, 请补充 (起点和终点请用数据流图 2-2 中的符号或名称, 数据流请用上文中的文字)。请在答题纸上直接作答, 不要重新画图。



图 2-1 0 层数据流图



2. 试卷总结

2.1 填空题

知识点在课本上几乎都能找到，考前一定要认真过一遍课本，建议用荧光笔标记每页出现的名词，方便查阅。

2.2 选择题

相比于填空题需要对课本知识点有一定理解。

2.3 简答题

不难，抄书就完事了。

2.4 应用题

1. 数据流图，填空：外部实体、数据存储、流；
2. 类图：给定背景和类图，填写图中类名；
3. 白盒测试
 - 计算环路复杂度
 - 写出独立路径
 - 写出测试用例

3. 复习建议

课本很重要，像软工这种完形填空考试答案课本上都有。

三、计算机网络

1. 题型

选择题&填空题

只记得我没写出来的几个点，感觉有点偏。

1. ADSL是什么复用
2. PCM: 采样、量化、编码
3. 海明码

简答题

1. 什么是协议？协议三要素？
2. 物理接口四个特性及具体实例
3. 数据链路层如何实现速度匹配和可靠传输？
4. 计算成功传输时间：两地距离1000m, 传输速率10Mbps, 数据量1500B, 传播速度 $2 \times 10^8 m/s$ 。
5. UDP和IP都是无连接的不可靠协议，能否直接用IP为应用层提供服务？为什么？
6. 电子邮件的4个协议是什么？作用分别是什么？

综合题

1. RIP协议
 - 简述RIP原理
 - R5与R1-R4相邻，给了R1-R4的路由表，计算R5的路由表。
2. TCP
 - 简述快重传，快恢复原理
 - 1-20次传输的拥塞窗口大小变化及原因分别是什么

2. 复习建议

可以结合谢希仁教材和李老师PPT复习，有时间做点王道考研的题。IP和TCP是重点，大题一般也会考。ps 应用层协议也要记下。

四、自然语言处理

1. 题型

参考: <https://github.com/VayneDuan>

简单说一下变化：

- 计算题

1. CNN计算: 1) 算卷积 2) padding后算卷积 (步长改变了, 我没看见被坑了) 3) 算average-pooling 4) 算RELU值
 2. 维特比算法: 1) 维特比变量 2) 给出隐状态序列
- 简答题记不得了, 难度不大, 就是抄书
 - 综合题2021年是3个, 也是抄书, 不过很多, 记得抓紧时间抄书。

今年NLP难度还行, 题量很恐怖, 抄书抄的手快断了, 有的同学都没写完。

五、机器学习

说在前面: 今年ML计算题改成了算法题, 所以更像文科了, 主要就是看懂西瓜书和PPT上的算法, 背下重要的知识点就行。

题目回忆可能与真实题目有出入, 大致意思是一样的。

1. 题型

简答题 (顺序不一定对)

1. 简述假设空间和版本空间。
2. 简述训练误差、测试误差、泛化误差。是否训练误差小, 泛化误差一定小?
3. 朴素贝叶斯解决了什么困难? 提出的假设是什么?
4. 对数几率回归解决的是回归问题还是二分类问题? 训练方法是什么?

算法题

1. 补充注释: k-means聚类, 补充几条语句的注释

输入: 样本集 $D = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$;
 聚类簇数 k .

过程:

- 1: 从 D 中随机选择 k 个样本作为初始均值向量 $\{\mu_1, \mu_2, \dots, \mu_k\}$
- 2: **repeat**
- 3: 令 $C_i = \emptyset$ ($1 \leq i \leq k$)
- 4: **for** $j = 1, \dots, m$ **do**
- 5: 计算样本 \mathbf{x}_j 与各均值向量 μ_i ($1 \leq i \leq k$) 的距离: $d_{ji} = \|\mathbf{x}_j - \mu_i\|_2$;
- 6: 根据距离最近的均值向量确定 \mathbf{x}_j 的簇标记: $\lambda_j = \arg \min_{i \in \{1, 2, \dots, k\}} d_{ji}$;
- 7: 将样本 \mathbf{x}_j 划入相应的簇: $C_{\lambda_j} = C_{\lambda_j} \cup \{\mathbf{x}_j\}$;
- 8: **end for**
- 9: **for** $i = 1, \dots, k$ **do**
- 10: 计算新均值向量: $\mu'_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x}$;
- 11: **if** $\mu'_i \neq \mu_i$ **then**
- 12: 将当前均值向量 μ_i 更新为 μ'_i
- 13: **else**
- 14: 保持当前均值向量不变
- 15: **end if**
- 16: **end for**
- 17: **until** 当前均值向量均未更新
- 18: **return** 簇划分结果

输出: 簇划分 $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$

上面这段代码中的中文部分。

2. 补充代码: 补充AdaBoost的代码

输入: 训练集 $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$;
 基学习算法 \mathcal{L} ;
 训练轮数 T .

过程:

- 1: $\mathcal{D}_1(\mathbf{x}) = 1/m$.
- 2: **for** $t = 1, 2, \dots, T$ **do**
- 3: $h_t = \mathcal{L}(D, \mathcal{D}_t)$;
- 4: $\epsilon_t = P_{\mathbf{x} \sim \mathcal{D}_t}(h_t(\mathbf{x}) \neq f(\mathbf{x}))$;
- 5: **if** $\epsilon_t > 0.5$ **then break**
- 6: $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$;
- 7:
$$\begin{aligned} \mathcal{D}_{t+1}(\mathbf{x}) &= \frac{\mathcal{D}_t(\mathbf{x})}{Z_t} \times \begin{cases} \exp(-\alpha_t), & \text{if } h_t(\mathbf{x}) = f(\mathbf{x}) \\ \exp(\alpha_t), & \text{if } h_t(\mathbf{x}) \neq f(\mathbf{x}) \end{cases} \\ &= \frac{\mathcal{D}_t(\mathbf{x}) \exp(-\alpha_t f(\mathbf{x}) h_t(\mathbf{x}))}{Z_t} \end{aligned}$$
- 8: **end for**

输出: $H(\mathbf{x}) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(\mathbf{x}) \right)$

综合题

1. SVM

- SVM分类还是回归?
- 不能正确分类的问题还能用基本型SVM解决吗?
- 上一问的解决办法: 我回答的是软间隔和核方法

2. 集成学习

- 根据图片判断集成学习是否提升性能

测试例1 测试例2 测试例3				测试例1 测试例2 测试例3				测试例1 测试例2 测试例3			
h_1	✓	✓	×	h_1	✓	✓	×	h_1	✓	×	×
h_2	×	✓	✓	h_2	✓	✓	×	h_2	×	✓	×
h_3	✓	×	✓	h_3	✓	✓	×	h_3	×	×	✓
集群	✓	✓	✓	集群	✓	✓	×	集群	×	×	×
(a) 集群提升性能				(b) 集群不起作用				(c) 集群起负作用			

- 简述影响集成学习性能的因素

$$E = \overline{E} - \overline{A}$$

- ▣ 这个漂亮的式子显示:个体学习器精确性越高、多样性越大,则集成效果越好。称为误差-分歧分解

- 枚举提升集成学习效果的方法

2. 复习建议

考前看下西瓜书+陈老师的PPT就行了,主要是概念要理解和记忆,尽量全面一点(防止考到没看过的概念)。像常考的点,比如:过拟合vs欠拟合,分类vs回归,监督学习vs非监督学习.....,最好先总结答案考试直接默写。