

## 2019 年（2017 级大三上考试）

### 编译原理期中：

#### 一、填空题（20 分，20 个空）：

词法分析器输入输出是什么，是什么文法，语法分析器输入输出是什么，是什么文法（6 分）

前端后端组成（5 分）：全知道就完事儿了，填了 5 个空

词法分析器的构造（5 分）：5 个部分

编写词法分析器的三种方法（2 分）：只填两个空

判定 LL(1) 文法的两个条件（2 分）

#### 二、大题（80 分，4 道）：

##### 1. 写正则表达式（送分题）：

（1）以 a 开头以 a 结尾的字符串的正则表达式

（2）以偶数结尾的正则表达式

##### 2. 写 FIRST 集和 FOLLOW 集（10 分）：送分题

##### 3. LL(1) 文法（25 分）：考了提取公因式、LL(1) 分析表和字符串匹配过程

##### 4. LR(1) 文法（35 分）：一共会有 18 个状态，不管你们考的是啥，要稳住

### 编译原理期末（哎，这是被球哥坑掉的一届，题太恶心了，祝好运）

#### 一、 填空题（32 空，32 分，哎~）

活动记录 7 个空

符号表 7 个空

循环优化的方法 4 个空

词法分析器和语法分析器的关系 3 个空

数组地址 4 个空（比如  $a[i]$  的地址含义乱七八糟、从  $a[i]$  到  $b[j]$  的花销，反正这里需要理解就完事儿了）

剩下的想不起来了，真的填空题太恶心了，希望学弟学妹们不要遇到这么恶心的题！！！！

#### 二、 大题（68 分）

##### 1. SLR 分析表，分析字符串（24 分）

##### 2. 消除左递归，写新的语法制导方案（十几分）

##### 3. 画 Display 表（十几分）

##### 4. 给了代码段，写引用相关信息。（十几分）

## OOAD 期中: (大题和简答题分值各占一半)

### 简答一道 5 分

1. Explain Analyze/Design、OOA/D (2019 考过)
2. OOA&D vs structured A&D (2019 考过)
3. What is EBP? (2019 考过)
4. UP 定义 (2019 考过)
5. UP 和 OOAD 的关系 (2019 考过)
6. use case model vs feature list (2019 考过)
7. use case 是否是 a good one (2019 考过)
8. How to create a domain model (2019 考过)
9. (大题)用例图和写文字用例 (2019 考过)
10. (大题)建 domain model, 以及用到什么技术 (确定名词短语 . . . . . ) (2019 考过)
11. OOA, OOD, OOP 的关联 (2019 考过)

往年还考过:

迭代过程和敏捷模型

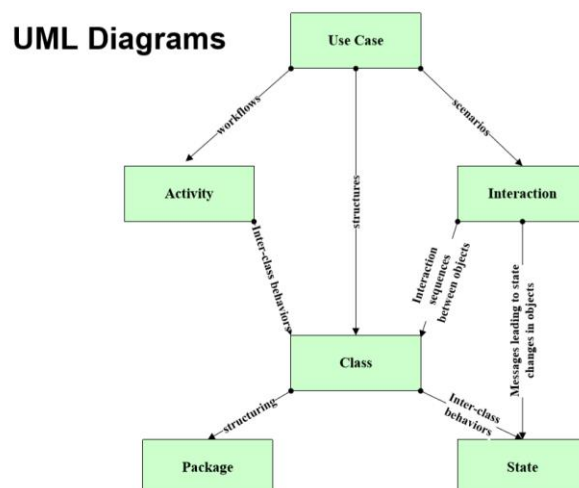
## OOAD 期末:

### 一、 简答题 (5 分一道)

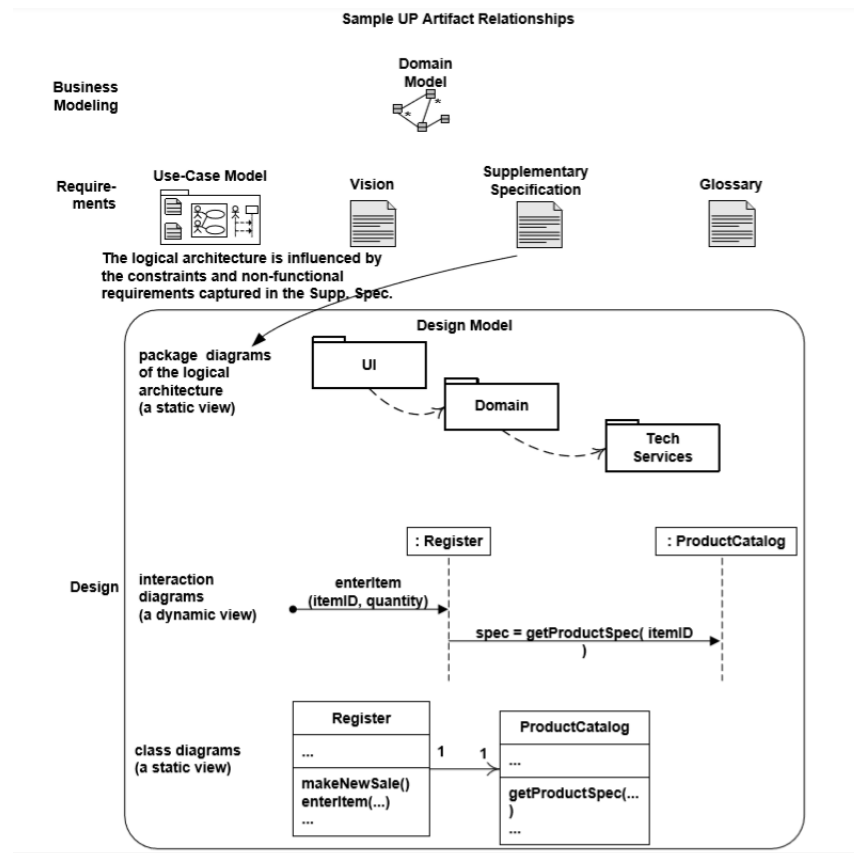
1. what are three main topics about OOAD course? 敏捷建模? ooa? ood?
2. In which diagram only object in UML and in which only class? 领域模型? 类图?
3. What is software design pattern?
4. Three ways of using UML
5. Under which conditions do we need 操作契约

### 二、 大题

- 1、解释这个图 (20 分)



- 2、Name five GRASP patterns and explain them with short text and UML
- 3、说出用了什么模型，explain the diagram and the idea behind the specific example



## 数据挖掘期末:

一、 选择题（30 分，每题两分）：（排序不分先后，不记得具体顺序）

1. 数据归一化计算
 
$$v' = \frac{v - \min_A}{\max_A - \min_A} (new\_max_A - new\_min_A) + new\_min_A$$
2. 处理噪声的方法（选项里有贝叶斯）
3. 分类算法特有的要求（过拟合）
4. 简单地把集合划分成几个子集，然后每个数据恰在其中一个，这是什么方法（聚类方法）。
5. 距离度量比 0 重要的二元属性（非对称的二元属性）
6. 下面几个选项那个是数据挖掘任务（离群点检测）
7. 数据挖掘要什么样的数据（合适的的数据）
8. 关于数据挖掘找不正确的一项（选项里有：不是必须用数据仓库，不是必须得请编外专家，信息增益越高效果越好（这是错的，选这个））
9. 决策树不包括以下哪种结点？（C）
  - A. 根结点
  - B. 内部节点
  - C. 外部节点
  - D. 叶节点

其他的问了一圈大家不记得了，但肯定是答案显而易见的。

## 二、 判断题（10 分，5 个）：

5 个小题判断它们是否为数据挖掘任务

其中有个数据仓库的问题（应该算是数据挖掘任务）

## 三、 填空题（10 分）：

1. K-means 计算聚类集合（4 分）

类似下图做法，但比这个还简单

### k-means 算法示例

序号	属性1	属性2
1	1	1
2	2	1
3	1	2
4	2	2
5	4	3
6	5	3
7	4	4
8	5	4

设 $n=8$ ,  $k=2$ ;

第一次迭代：随机  
选择序号1和3作为  
初始点。

找到离二点最近的  
对象，产生二个簇  
{1,2}和{3,4,5,6,7,8}

均值点分别为  
(1.5,1) (3.5,3)

序号	属性1	属性2
1	1	1
2	2	1
3	1	2
4	2	2
5	4	3
6	5	3
7	4	4
8	5	4

第二次迭代：通过平均值调整对象  
所在的簇，重新聚类。按离平均值  
点 (1.5,1) 和 (3.5,3) 最近原则重  
新分配，得到新簇 (1,2,3,4)，  
(5,6,7,8) 计算新的平均值点  
(1.5,1.5)，(4.5,3.5)

第三次迭代：将所有点按离平均值  
点 (1.5,1.5) (4.5,3.5) 最近原则  
重新聚类调整簇，簇依然为  
(1,2,3,4)，(5,6,7,8)，没发生  
重新分配，程序结束。

2. 数据挖掘标准流程，六个空（6 分）

## 四、 计算(15 分)：

Apriori 算法和 FP-growth 算法求频繁项集，并找出关联规则。

## 五、 简答（3 个，共 35 分）：

1. 假如你是一个高校的数据分析师，如何应用 CRISP-DM，针对聚类、分类、关联分析、异常检测等技术，做对学校有用的数据挖掘工作；（12 分）

2. 为什么要进行数据预处理？有哪些方法，举例说明。（12 分）

3. 银行分析你的信用卡行为模式，发现你在家具店有消费，向你推荐家具贷款，这样算不算侵犯隐私，有什么不侵犯隐私并且还能让商家进行数据挖掘的解决办法吗？（11 分）

## 算法期末：（努力了，只记得这点了，这届考的略硬核）

### 一、 选择题（20 分，10 道，一道 2 分）

- (1) 算法的五个特性
- (2) 打印所有素数违背了五大特性中的哪一个（有穷性）
- (3) 自底向上的算法（动态规划）
- (4) 什么算法不能用分治法解决
- (5) 回溯法用（深度优先）
- (6) c 个学生，记录它们的成绩的和 sum，选语句  
sum->sum+t  
c->c+1

### 二、 判断题（20 分，10 道）就记得这几个，其他几个应该不难


- (1) 动态规划最优化原理和子问题重叠？
- (2) 递归设计总是需要初始值？
- (3) Best, worst, average 时间，其中 worst 具有最大的操作性和实用价值？

### 三、 填空题（20 分，10 道）

- (1) 给出了一个数列，求快排第一趟排序
- (2) 算法分析时间复杂度和（空间）复杂度
- (3) 折半查找的最坏情况的时间复杂度。
- (4) 求时间复杂度  
For(int i=0;i<=n;i++)  
For(int j=0;j<=2n;j++)
- (5) 贪心算法：最优子结构和贪心策略
- (6)  $T(n)=aT(n/b)+cn$  这个公式，给了一个公式，求  $O()$ 。
- (7) Prim、快排、哈夫曼算法中，哪个不属于贪心算法（快排）
- (8) 动态规划法存储 subsolution 的目的是什么。
- (9)  $O$ 、 $\Theta$ 、 $\Omega$  哪个用来作为算法确切的标准？？（好像是这么回事）
- (10) 下面这张图出了题

A, B 两个算法，一个是  $2T(n/2)+n$ ，一个是  $aT(n/4)+n$ ，A 小于 B，求 a 的最大值。需要会推。

$$T(n) = \begin{cases} c & n = 1 \\ aT(n/b) + cn^k & n > 1 \end{cases}$$


$$T(n) = \begin{cases} O(n^{\log_b a}) & a > b^k \\ O(n^k \log_b n) & a = b^k \\ O(n^k) & a < b^k \end{cases}$$

### 四、 求解（25 分，5 道）

- (1) 办活动，给出了开始时间和结束时间，要求有最大兼容量（区间调度

的变体），写出子集。

(2) 01 背包问题回溯法，画解空间树。

(3) 有 7 个小区，给出了坐标，现在要建医疗诊所，求建在哪里（邮局问题变体，是中位数）。

(4) 任务分配问题，画分支界限树。

(5) TSP **求路径和最小距离，给了一个图**

## 五、 算法设计题（15 分，2 道）

(1)  $n$  个人在一个服务窗口，给出了每个人等待的时间，求总的最小等待时间。（可以先排序，再计算）

(2) 展览馆办活动，给出了开始时间、结束时间、价值，动态规划求总价值最大（带权区间调度的变体）

## XML 期末

### 一、 选择（12 道，24 分）很简单！

1. xml 不是一种编程语言
2. xml 的处理指令
3. SOAP 的 error 的 subelement
4. WSDL 是 IDL
5. 内部声明 DTD 文件，`<!DOC root []>`
6. 下列选项种哪个是 well-formed xml
7. 下列哪个不能被此 DTD 通过。
8. SAX 节点的处理的顺序

### 二、 大题（76 分）

1. 给了 xml 写 schema

注意`<color value="red"/>`空元素的写法和`<description>kajdhfah <b>afgag</b></description>`混合元素的写法。

2. 给了 html 写 xsl
3. 给了 xml 写 XQuery 的结果，for return, let return
4. 写 JDK1.6 SOAP 的 request 和 response, WSDL 的 message 和 portType

## 计网期末（应用层不要忽略，要会核心的东西）

### 一、 填空（20 分）不难，是很基本的东西

应用层考过邮件传输协议、服务器和客户端之间传输的协议。

### 二、 选择题（20 分）不难，是比较基本的东西

### 三、 判断题（20 分）有一些比较细节的东西

Tcp 是字节流还是消息流？

UDP 有没有拥塞控制

### 四、 解答题（40 分）

1. DNS, 访问 [www.bupt.edu.cn](http://www.bupt.edu.cn), 如果本地服务器中没有此地址, 写出递归查询的过程, 画出图 (8 分)
2. 描述 TCP 如何保证传输安全的? 确认重传、数据校验、数据分片、流量控制、拥塞控制。
3. 写出 link state routing 的过程 这道题2018年期末考试题里出现过。
4. OSPF 和 BGP 的区别~
5. 下面这道作业题

3. 在一个 Ubuntu Linux 服务器中, 分别使用 route 命令和 ip route 命令打印出该服务器的路由表如下: ↵

```
abc@ubuntu:~$ route -n
Kernel IP routing table
Destination      Gateway         Genmask       Flags Metric Ref    Use Iface
0.0.0.0          192.168.25.101 0.0.0.0       UG        0      0          0 eth2
135.45.56.0      192.168.150.148 255.255.252.0 UG        0      0          0 eth1
135.45.60.0      192.168.152.129 255.255.252.0 UG        0      0          0 eth0
192.53.40.0      192.168.140.136 255.255.254.0 UG        0      0          0 eth3
192.168.25.0     0.0.0.0        255.255.255.0 U         0      0          0 eth2
192.168.140.0    0.0.0.0        255.255.255.0 U         0      0          0 eth3
192.168.150.0    0.0.0.0        255.255.255.0 U         0      0          0 eth1
192.168.152.0    0.0.0.0        255.255.255.0 U         0      0          0 eth0

abc@ubuntu:~$ ip route
default via 192.168.25.101 dev eth2
135.45.56.0/22 via 192.168.150.148 dev eth1
135.45.60.0/22 via 192.168.152.129 dev eth0
192.53.40.0/23 via 192.168.140.136 dev eth3
192.168.25.0/24 dev eth2 proto kernel scope link src 192.168.25.134
192.168.140.0/24 dev eth3 proto kernel scope link src 192.168.140.131
192.168.150.0/24 dev eth1 proto kernel scope link src 192.168.150.134
192.168.152.0/24 dev eth0 proto kernel scope link src 192.168.152.130
abc@ubuntu:~$
```

## 6. 还有这道作业题

1. 在熟悉以太网帧 (Frame) 结构、IP 分组 (IP Packet) 结构、TCP 报文 (TCP Segment) 结构的基础上, 针对下表列出的六条以太网帧进行分析: ↵
  - 1) 下列哪三条以太网帧中包含了一个 TCP 连接建立过程的三次握手过程? TCP 连接的服务器端和客户端选定的初始序列号 (ISN: Initial Sequence Number) 各是多少 (以 16 进制表示)? 三次握手过程中的第三条 TCP 报文的序列号 (Sequence Number) 是多少 (以 16 进制表示)? ↵
  - 2) 请写出这条 TCP 连接的服务器端的 IP 地址、客户端的 IP 地址。IP 地址以点分十进制表示。↵
  - 3) 这条 TCP 连接的客户端、服务器的端口号各是多少? ↵
  - 4) 第 5 个以太网帧中 TCP 报文的序列号是多少 (以 16 进制表示)? 携带的数据 (Payload) 的长度是多少字节? 设发出第 5 个以太网帧的一方为主机 A, 主机 A 的 TCP 协议在发出第五个帧中的 TCP 报文后接着发送后续数据, 这个接下来发出的 TCP 报文的序列号是多少 (以 16 进制表示)? ↵

Index	Frame															
1	ec	6c	9f	05	3e	5c	00	0e	c6	c2	79	48	08	00	45	00
	00	34	55	ac	40	00	80	06	00	00	ac	10	01	da	6a	0b
	d1	1d	c5	05	00	50	ee	8d	d6	c5	00	00	00	80	02	
	fa	f0	e9	39	00	00	02	04	05	b4	01	03	03	08	01	01
	04	02														

## 软件需求工程期末

(don't worry, 开卷考试, 基本都在 ppt 上)

选择题需要注意一道题: 软件需求获取的成果是什么 A. 录音和记录 B. 软件需求说明书 (最好查一下进行确认)

大题是分析酒店系统 (仅限前台人员使用) 的用例图、还有用例描述

总的来说很简单, 这门课想考 90 分以上要在实验上多下功夫, 多跟老师沟通, 混脸熟。

# 软件工程期末（我们这届没有考填空题）

复习的时候虽然 ppt 很多，but don't worry，ppt 第二部分是重点，其他地方好好读一读，基本会出在选择题和判断题

## 一、 选择题

没那么变态比较简单啦，老师真的没有为难我们。

## 二、 判断题

也很正常，不难。有一道印象深一点：软件生命周期是从软件开发到开发结束（错误的）

## 三、 大题

1. 给了一个电影售票系统，（1）画周境图（2）第一层数据流图（3）第二层数据流图（4）用例图。

2. B/S 结构（客户端，网页服务器，交易业务服务器，数据服务器）

（1） 分析用户多时，哪些服务器比较忙。

（2） 交易数多时，哪些服务器比较忙。

（3） 数据库每月统计时，哪些服务器比较忙。

（4） C/S 结构分层的优点

质量属性	对质量的影响
可使用性	每层的服务器可以部署多个服务器，互为备份，因此，一个服务器出现故障，其它服务器仍可使用。虽然会降低服务性能，但仍可用。
故障处理	如果客户端与服务器的通信失败，Web和应用服务器可以实现透明的失效备援。因此客户端的请求可以重新定位到运行着的备份服务器上，而客户并不知晓。
可修改性	分割增强了可修改性。表现层、业务和数据管理逻辑清晰地得到了封装。每个层面都有其内部逻辑，修改不会影响其它层。
性能	这种结构的性能已经得到了证明。关键的问题是要考虑每个服务器支持的并发线程数量、各层之间的连接速度、以及数据传递的速度。对于分布式系统来说，降低了为完成每个请求所需要的层与层之间的调用时间。
可伸缩性	各层中的服务器可以有备份，多个服务运行在同一个或多个不同服务器上，体系结构的规模可以得到很好地提升。在实际中，数据管理层往往会成为系统能力的瓶颈。

3. 语句覆盖，分支覆盖，MC/DC 覆盖 填表测试：

**Table 6-1:**

```
x = 3.0;
y = 3.0;
if (AA && BB)
{
    z = x + y;
    if (CC)
        { z = y + x/y;}
    else
        { y = x + y}
}
z = z + x/y;
```



答案:

The need of coverage test (测试要求)	Test-cases (测试用例)					
	Inputs			Outputs expected		
	AA	BB	CC	x	y	z
Statement coverage = 100% (语句覆盖达 100%)	1	1	1			
	1	1	0			
Branch coverage or Decision-Decision Path(DDP) = 100% (分支覆盖达 100%)	1	1	1			
	1	1	0			
	0	0	1			
MC/DC = 100%	1	1	1			
	1	1	0			
	0	1	1			
	1	0	0			

## 并行计算期末

这门课，听话，如果有考试的话不要乱选，考试很恶心。考的全是卢姥爷画的重点，但是如何考就比较出乎意料了，相比之下不如选数值计算，人家只有一个大实验，给分还基本 90+，并行计算学姐我多交了一个实验报告才勉强 88 分。

### 一、 填空题（10 个）

加速比公式、并行计算的条件、桶排序成功的条件、1974 年（好像是这个年份？）  
\_\_\_\_\_是并行计算开始流行的一年、并行计算几个功能、并行计算 3 种环境。

### 二、 选择题

好几道题是给了代码让选择输出，这个考阻塞和非阻塞的功能，一定要知道它的原理。

考了加速比的极限

其他的就不记得了，太难了。

### 三、 判断题

太难了，完全不记得，就记得第一道题还是考了加速比的极限。

### 四、 大题（很恶心，复习的时候并不知道要这样考）

写奇偶交换的原理和复杂度证明

写  $x^3$  个处理器处理  $n \times n$  矩阵的乘法的原理和复杂度说明

写流水线处理某问题的伪代码  
写出 moore 算法的原理和伪代码  
伪随机数生成器（模除余）的证明。

**嵌入式没有考试！**

**通软虽然实验难，但是考试很简单，雷老师很友善。**

学姐我没选这个课，但是据说考的很基础，大家都简单。  
sdl 图 如何定义数据结构 和 怎么画一个 tcp 的握手成功流程。

**Python 如果不需要修学分，就不要闲的非要参加这个考试了，比较 emmm，大家都懂。。。。**

**祝大家好运哦！！！！**