2016年3月软理 方向复试笔试

编译原理：

1. 写一个NFA，L={w|w属于{a,b}\*}。前5个字符至少一个aa子串，NFA状态不多于8个。
2. 写出一个正规式，L={w|w属于{a,b}\*}。当以a结尾时串个数为奇数。|、\*、连接符号不超过10个。
3. L={w|w属于{a,b}\*}。写出上下文无关文法，使得生成不对称串，非终结符个数小于等于3。
4. 活跃变量分析、到达定值分析、UD链分析。
5. IR（0）语法、SLR（1）语法

数据库系统

1. 简述事务的ACID特性
2. 简述数据库的故障恢复过程
3. 为什么不能在视图上进行增删改操作？
4. 什么是主索引？主索引可以是顺序存储，可以是B+树，为什么不能是hash存储？
5. 数据库并发控制的原则是什么？基于该原则出现了哪些并发控制机制？

2015年3月软理 方向复试笔试

数据库（共5道简答题，每题10分）：

1关系数据库系统的核心技术是什么

2简述事务的ACID特性

3请描述一下数据库的故障恢复过程

4数据库为什么要进行查询优化

5数据库中为什么要进行并发控制？都有哪些技术可以来完成并发控制？

编译原理（共3道大题，第一道15分，第二道24分，第三道8分，总分应该50分） 一（2道小题，第一道5分，第二道10分）

1写一个NFA，{a,b}其中不得出现连续2个b，NFA状态不多于6个。

2见下图（编译原理第一大题，基本类似），写每个语言的二义文法和无二义文法各一个，总共2个语言。

二（6道小题，都是基于LR分析的）

其中有SLR(1)填表和项目集，然后有关于LR(1)和LALR(1)的内容，最后是发现错误LALR

（1）和SLR（1）谁更早，举例说明。

具体可以参照《编译原理》张素琴那本书第7章练习题中第6、14、15题。

三（4道小题，关于代码优化部分）

第一个是填表，第二个第三个是关于ud链和du链的，第四个应该是填空。图形见下图（编译原理第三大题，类似）。

今年编译原理没有考DFA和NFA的转换，也没有考LL分析，所以编译原理的复习一定要全，看往年的贴子你就会发现每年题都不一样，考试的重点也不一样，所以经验贴意义不大。

清华CS复试2013笔试编译原理回忆

一. 给定一个文法

S -> S$S | a | b | c

问：

（1）将该文法转化为非二义性的上下文无关文法

（2）给出文法表示语言的正则表示

（3）给出文法对应的有限自动机

（第二题和第三题用的是同一个文法，具体题目的顺序记不清了，请后来的同学补充）

二.给定一个文法

S -> A

S -> ε

A -> a

A -> (S)S

并附文法LR（0）图一张（这个可以自己动手画一画）

问：

（1）求出各产生式右部符号的FIRST集以及产生式左部符号的FOLLOW集

（2）使用一问的结果填写递归下降分析表（一个LL（1）表）

（3）证明该文法是一个LL（1）文法

（4）看图指出图中哪些状态存在冲突，指出冲突的类型以此证明该文法不是LR（0）文法

（5）证明该文法是SLR（0）文法

（6）填写该文法的LR分析表

（7）'(a))'不是该文法的句子，请问使用LR分析时在哪个状态出错？

四.第四题是真的记不清了，这道题是关于DISPLAY表的，一共5分。

题目大概是这样的：给定一个类似PASCAL的程序（程序中允许嵌套定义函数）

然后给出一个DISPLAY表，问题是求DISPLAY表中第18、19、22、23行各应该填什么

这几行对应RA（返回地址）、SL（似乎叫静态链）、DL（动态链），然后把它们的内容补充完整就行了。

（DISPLAY表我不太熟悉，考试的时候几乎是蒙的，考题实际上很简单，明年考编译的同学可以多花点时间做点DISPLAY表方面的题）

2012清华软件理论方向笔试试题回忆版

每部分50分，共100分

数据库部分

一、数据库管理系统（DBMS）的核心技术是什么？（15分）

二、数据库中可以定义视图，为什么不能在视图上对数据库进行插入，修改，删除操作？（20分）（表述可能有出入，大致意思如此）

三、什么是主索引？主索引可以是顺序存储，可以是B+树，为什么不能是hash存储？（15分）

编译原理部分

一、画一个DFA，不超过5个状态，表示语言L(G)={w|w上含有且仅含有奇数个ab字串}，字母表{a,b}（6分）

二、写正规式，表示语言L(G)={w||w|>=2且w中至少含有两个位置不同的字符}，字母表{a,b}（表述可能不准确，欢迎指正）（6分）

三、写一个上下文无关文法，表示字母表{a,b}上的其中不含ab串的字符串，非终结符不超过4个。（6分）

四、给一个LR（1）的DFA，有三问：第一问让补充完整DFA；第二问问项目规范集中存在什么冲突；第三问让写完整的LR（1）分析表（18分）

五、语义处理题，描述很长，让你按照描述写（1）E->E(与非)E （2）repeat S until E

的语义规则，题目很麻烦，教材上也没有相似类型的题目，我是照葫芦画瓢仿写的，也不知道对不对。（14分）

另外推荐下参考书目，都是清华自己出的教材：

数据库用冯建华，郝晓龙，周立柱编著的《数据库系统设计与原理（第二版）》，清华大学出版社

编译原理用张素琴，吕映芝，蒋维度，戴桂兰编著的《编译原理（第二版）》，清华大学出版社

2011复试笔试各科题目

数据库：（一共两大题。第一大题30分。第二大题20分）

一1、你觉得数据库管理系统的核心技术是什么

2、事务的ACID特性指什么

3、数据库冲突处理的原则是什么。根据这个原则。数据库有哪些冲突处理的方法。

二、数据库以B+树结构存储。分为索引页与叶级页。两级索引。索引页有8000B可以用来存储索引项。每个索引项80B

第一级索引页有两个索引项。问整个数据库的最大存储空间与最小存储空间是多少。

索引那个题还有个条件是索引页有8000B可以用来存储索引项（就不用转换KB的1024了）

编译原理

1.画一个DFA，要求描述在字母表{a,b}上含有子串aa但不含有子串aaa的字符串，状态数不超过6个

2.写一个正规式，要求描述在字母表{a,b}上不含有bb子串的字符串，正规式长度不超过8，长度包括连接、或、闭包，但不包括括号

3.写文法，要求是{a^nb^mc^k | n<=k, m+n >= k}( n,m,k 都是幂 )

剩下的题记不太清了，大体上

1.给出LR（0）的DFA，考 LR(0) SLR(0) 以及找出冲突状态，并探究哪些状态用某种方法可以消除冲突

2.给出LR（1）的DFA，求LALR(1)对应的同芯项目集，以及对一个给定字符串进行规约时使用LR（1）和LALR（1）哪个可以更快发现错误

3.一个流图的问题，考程序点（可能是这个概念，我没复习到这块）和DU链，这题我放弃了

2010清华计算机系软件理论方向笔试题简忆

一、数据库

1、什么是数据库，数据库管理系统，数据库系统

2、简述什么是数据库事务的ACID特性

3、一个表student(name,age,sex)，求不同姓名的学生的个数（表的属性肯能不太准确）

4、什么是主索引，主索引可以是顺序存储，可以是B+树，为什么不能是hash存储（措辞可能不太准确）

第16 / 17页

5、简述当前数据库研究的五大热点

二、编译原理

1、字符表为{a,b}，求开头和结尾相同，长度大于等于2的正规式，要求只能用连接，闭包和另外i一个(忘了)，正规式长度不能超过10，不计括号

2、字符表为{a,b}，求前5个字符包含ab字符串，长度可能小于5，的含ε的NFA,要求状态不超过6个。