

Architecture

基本上都是书上面的原题（张晨曦，计算机系统结构），稍稍改了改。选择题部分：

1. 一个部件执行时间是0.5，加速比为20，那么系统整体的加速比为多少
2. 按照Flynn分类法，多处理机器属于哪一种？（SISD,SIMD,MISD,MIMD）。
3. 关于RISC体系结构那种说法错误？（指令种类多，指令简单统一，指令执行时间快，还有一个忘了）
4. 哪一种方式对于组相连映像Cache的失效率没有影响？（增大内存大小、增大Cache大小、增加组相连度， $\#@\# \text{ ¥ } @\# \text{ ¥ }$ ）
5. 哪一种超标量处理机的资源使用率最大？（支持多线程技术的超标量处理机、支持粗粒度多线程的超标量处理机、支持细粒度多线程的超标量处理机、支持同时多线程的超标量处理机）
6. 关于直接映像Cache的索引和标识位数的计算。只记得这么多了 计算题部分：
7. P24 1.6
8. P109 3.15
9. P220 5.11
10. P250 6.8
11. P110 3.17 综合题部分：
12. P150 例题4.8 （循环展开3次）
13. P288 例题8.3 （SC指令不占用总线事务）

1. 某一功能部件，加速比为10倍。占40% 整体性能提升多少？ 1.45 1.56 1.28 1.36
2. MIPS的5段流水线，问有几个IR（指令寄存器）？ 1 2 3 4 5
3. 直接映象cache一般使用那种替换机制 随机 FIFO LRU 前三个都不是
4. 一个数据通道 选择时间为 T_s ，传输一个字节需要 T_d ，问选择通道一次传送 n 个字节。问该通道的最大流量
5. 有一延时分支，问在何时该分支的性能最佳？从失败取指令 从成功取指令 从前取指令 第四个选项忘了
6. 前瞻执行有哪四个步骤。给出了6-8个备选项
7. Tomasulo 的变换寄存器的名字能消除哪些冲突？多选 RAR WAW RAW WAR
8. 一个六段流水线 先执行一个由 L 个互不冲突的指令的一段程序，都执行完之后 接着执行由 $2L$ 个互不冲突的指令的程序 问共需多少时钟周期 $3L+5$ $3L+6$ $3L+10$ $6L-6$
9. 用监听（Snoopy）机制的多处理机 cache一致性方案，通常采用的是？多处理机系统 SIMD 松散指令 总线互联 交叉开关
10. 多核处理机属于？用？方式通信 SISD SIMD MISD MIMD SMP NUMA 网络

第二题 40分 1. 一个用写回法的直接cache 索引2ns 标识1ns 处理机访问cache4ns TLB 2ns 物理cache 读写命中的时候，时间各至少是多少 虚拟cache 读写命中的时候，时间各至少是多少

2. 一个静态的5段流水线 乘法指令用1,2,3,5 段 加法指令用1,4,5 段 每段的时间 $1 \rightarrow t_1$ $2 \rightarrow t_2$ $3 \rightarrow 2t_1$ $4 \rightarrow t_1$ $5 \rightarrow t_1$ 计算 $a_1b_1+a_2b_2+a_3b_3+a_4b_4$ 画图, 求时间, 求与无流水线比较时的加速比

3. 一个15MHZ的cpu, 10MIPS 30%的访存 多一次存储器 5%的访存 多两次存储器访问 一次存储器访问 1个周期
1. 求有效CPI 2. 若改成30MHZ的cpu, 求性能改进。此时访存时间不变, 就是变成2个周期。

4. 分析WAR WAW的冲突, 要多用几个寄存器才能消除相关 LD R1 0(R2) ADD R3 R4 R1 ADD R4 R5 R6 ADD R7 R4 R8 ADD R8 R9 R10 ADD R11 R8 R12 ADD R12 R13 R14 ST 0(R15) R12

5. 24维德向量 加法2拍 乘法3拍 求(A+B)*C 求流过时间 一个周期50ns和MFLOPS

第三题 30分

1. 一个超标量处理机, 一次能流出两条指令 一条整形, 一条浮点 LD F0 0(R1) ADD F4 F0 F2 SD F4 0(R1) DADDIU R1, R1, #8 JNE R1, R2, LOOP 其中两条指令之间无延时, 只有浮点运算和浮点store之间有一个延时 跳转指令后有一个延时槽 求一种新的方案, 执行时间最短

2. 填写两张表格, 是关于多处理机中cache一致的两种实现方法的, 和书上一样 (类似) 的两张表吧

3. 用LL/SC实现fetch_and_increase_one 设n个处理器要执行该指令 并计算总线 最多和最少的情况下要用几次总线操作。假设SC处理不成功是不算总线操作

一、正则表达式==》NFA==》DFA

二、First、follow 集合, LR(1) 语法, LR(1) 表, 判断LR(1) 语法

三、写出简单while循环的IR形式

四、寄存器分配题 (此类题目有三种类型: 简单的不带合并的算法, 带合并的算法, 带预着色的算法-此类最复杂)

五、划分程序基本块、画出变量冲突图

1 简答题

图形和图像的区别、定义

光栅、光栅显示器、光栅图形、光栅化、光栅装换的概念解释

2 选择题

GKS是什么意思, 以及和这个一类的缩写所代表的含义。选择题和网上搜到的选择题没有多大出入, 感觉好多一样

判断题: 数值微分法、bresenham方法, 判断Bezier曲线是否是对称的

3 计算题:

区间扫描线算法基本原理 (讲算法就好)

字符串裁剪三种类型 (几乎必考之送分题)

旋转、加粗、斜体是怎么做的 (ppt、书上原题)

二维连续旋转变换有一下公式成立： $R(\theta_1) * R(\theta_2) = R(\theta_1 + \theta_2)$ 。只要用旋转矩阵证明就好

简述曲线的六种连续性： $C_0, C_1, C_2, G_0, G_1, G_2$

简述/伪代码描述 消隐算法Z-buffer，以及优劣（消隐算法总共6个，去年考了4叉树消隐，今年z-buffer，记一下，应该是必考点）

NURBS的定义、权重、局部性、凸包、变差缩减性说明（应该和B样条是交替考，都不深，主要考理解和概念）

集中复习：光栅图形学、曲线曲面、图形变换即可，其他了解概念。

一学期不上，10天可以复习的很好。（虽然不知道自己考得怎样，我只用了2、3天）

2018

五道大题

一、正则表达式相关

1 给定正则表达式，说明意思

2 给定要求，写正则表达式：匹配url： $(https|http|ftp)://((\w|-)+(\w|-)/[a-zA-Z0-9-_\&%?])$

二、维特比算法 标注给定句子中词的词性：人民(n) 生活(n/v)水平(n)得到(n/v)一定(d)提高(n/v)

三、pcfg 求出给定的句子的结构：给定一堆文法以及一个句子（i saw a dog with a telescope） 1》pcfg过程和结果状态图，2》结构树，3》概率

四、代码查重：从nlp的角度：1》可能有哪些情况，2》怎么解决

五、针对淘宝卖家刷评论、虚假评论的解决方案，nlp角度：1》可能有哪些情况，2》怎么解决

2017

自然语言处理2016年期末考题

一、Python字符串处理（10分）

已知text = "hello world start president"

1. 使用split将text分割成词组sent，并找出s开头的词

2. 找出sent中长度大于5的词

3. 取每个词的第三个字母，并连接成字符串

二、正则表达式（15分）

1. 识别有1~3位小数的正实数的正则表达式

2. 识别开头为大写的英文单词

3. 识别8~10位的密码，第一位必须为字母，后面由字母、数字、下划线组成

4. 写出含义并举例： $\wedge-?\d+(\d*)?\$$

5. 写出含义并举例： $\wedge \backslash w+([-+]\w+)@ \backslash w+([-]\w+). \backslash w+([-]\w+)^* \$$

6. 写出含义并举例： $\wedge [a-zA-Z0-9]^* \$$

三、词性标注（15分）

已知

```
from nltk.corpus import brown
```

1. 找出brown.tagged_words()中出现过1000次以上的词

2. 找出brown.tagged_words()中出现在动词后面最常出现的10种词性

3. 分割出合理的训练集与测试集，利用默认标注器，一元标注器与N元标注器，构造组合标注器对测试集进行评估

给出的提示：`nltk.FreqDist / nltk.tagged_sents / nltk.DefaultTagger / nltk.UnigramTagger / nltk.BigramTagger`

四、分类（15分）

已知

```
from nltk.corpus import movie_reviews
```

```
documents = [(list(movie_reviews.words(fileid)), category)
```

```
for category in movie_reviews.categories()
```

```
for fileid in movie_reviews.fileids(category)]
```

利用朴素贝叶斯分类器对文档进行训练与分类

提示：类别只有pos与neg，`nltk.NaiveBayesClassifier / nltk.classify.accuracy`

五、文法分析（20分）

给定句子“Mary saw a dog in the street”，以及文法：

$S \rightarrow NP VP$

$VP \rightarrow V NP \mid V NP PP$

$NP \rightarrow Mary \mid Det N \mid Det N PP$

$P \rightarrow in$

$N \rightarrow dog \mid street$

$Det \rightarrow a \mid the$

$V \rightarrow saw$

$PP \rightarrow P NP$

1. 写出线图chart算法的最终状态图

2. 画出语法分析树

六、语言建模（25分）

1. 做高考语文的文言文翻译与作文题

2. 对高考语文的文言文翻译与作文题进行自动批改

\1. list two typical PR learning paradigms and briefly explain the basic ideas. Give two applications and the give potential algorithms to solving the learning problems, respectively.

\2. explain the objective function of support vector machine for the linearly non-separable case and list the primal and dual formulations for optimization .

\3. show the basic ideas for linear regression and a generalized linear model and give a real application, respectively.

\4. give a typical pattern recognition and machine learning inference system based on inductive learning, e.g. the SVM algorithm.

\5. Consider two categories, each of which is described by a d -dimensional Gaussian having the same (but arbitrary) covariance, arbitrary means and arbitrary priors.

(1) give the minimum-error decision boundary.

(2) express the linear classifier with the parameters w and x_0 in terms of the variables given.

\1. 根据图示解释 Bresenham 画线算法。教材 P169

\2. 画图解释字符串裁剪的三种精度。PPT 第四章 P98

\3. 根据图示解释线段裁剪的中点分割算法。PPT 第四章 P83

\4. 圆锥曲线的 NURBS 表示，写出权因子分别等于哪些值时对应的曲线形状是什么。教材 P323

（由于试卷印刷出了意外，此题在考试时作废，换成了这样一道题：阐述计算机图形学、计算机视觉和图形处理之间的联系和区别）

\5. 选择和判断题，各 5 题。比较水，记下来的有：

显示器最常用的颜色方案

NURBS 曲线的优点

帧缓冲区大小计算

Bezier 曲线有对称性

齐次坐标可以表示无穷远点

\6. B 样条曲线上参数区间 $[t_i, t_{i+1}]$ 的一段与哪些控制顶点有关？在哪个凸包内？移动 P_i 影响哪个区间？教材 P309-310

\7. 给出二次 B 样条曲线的矩阵形式，计算端点、导矢、二阶导矢，然后作图并说明几何意义。教材 P312

\8. 写出一次、二次 Bezier 曲线的代数形式和矩阵形式，并描述它们的形状；用 de Casteljau 做出 $n = 3, t = 0.25$ 的曲线上的一点。教材 P303-305

\9. 写出绕 AB 轴做旋转的变换矩阵。教材 P371

\10. 根据图示写出平面对直线段的遮挡判断算法。PPT 第九章 P7

$((a|b)b)^* \rightarrow \text{NFA} \rightarrow \text{DFA}$

第二题：

文法

$S \rightarrow TB$

$T \rightarrow Ba|\epsilon$

$B \rightarrow Db|eT|\epsilon$

$D \rightarrow d|\epsilon$

(ϵ 代表空，打不出那个符号)

然后判断是否LL (1) 文法

第三题：

$a[i+1]=a[i]+1;$

转化为IR树，用Maximal Munch算法求指令集

第四题：

冲突图求寄存器分配（合并，染色）

第五题：

一坨代码执行顺序图让你求冲突图（就是画那个in out表格）

\1. 什么是计算机图形学，主要研究内容是什么

\2. Coons曲面四条边的曲线表示

\3. 中点画线算法的基本原理，快速方法

\4. 扫描线填充算法基本原理，给了书上的图和新边表，结合图说明活性边表的迭代更新过程

\5. Bezier曲线的分割递推de Casteljau算法，作图画出曲线上一点， $n=3$ ， $t=1/2$

\6. 两条曲线又几种连续方式，详细说明

\7. 多边形剪裁的Sutherland算法的基本原理，给了书上的图，画出每一步的结果

\8. 3次B样条端点的位置矢量，一阶导矢，二阶导矢的表示和几何含义（给了矩阵表示，只推导即可）

\9. 错切变幻，在x方向系数b，y方向系数d的错切变换，最后的参数表示结果和变换矩阵

\10. z-buffer算法说明（伪代码文字均可），及优缺点