

哈尔滨工业大学 2020 年春季学期

《信息内容安全》期末考试试题 (C)

一 简答题 (30 分) :

1: (6 分) 请用概要的文字简述如下问题, ①什么是国家的网络空间主权? ②从信息内容安全的主旨出发, 如何保障我国的网络空间主权?

2: (6 分) KMP 算法和 BM 算法都是经典的单模式匹配算法, 这两种算法中, ①属于前缀搜索算法的是? ②属于后缀搜索算法的是? ③如分别使用 KMP 算法和 BM 算法对如下的文本串与模式串进行匹配, 需要进行的子串对齐 (attempts) 次数和字符比较 (character comparisons) 次数分别是?

文本串: GCATCGCAGATACGATATATACG

模式串: ATATA

3: (6 分) ①影响网络数据被动捕获系统性能的关键因素有哪些? ②传统的被动报文捕获方法中数据由网卡到用户空间需要经过哪几次数据拷贝? ③零拷贝技术的核心思想是什么?

4: (6 分) 相对于 Web1.0, Web2.0 网络信息获取的难度更大, 主要难点有哪些?

5: (6 分) 欧盟通用数据保护条例 (GDPR) 2018 年 5 月 25 日正式生效, 并扩展了个人数据的定义, ①在 GDPR 中定义的个人数据包括哪些? ②大数据背景下, 个人数据的隐私保护问题遇到的技术挑战有哪些?

二 计算题 (20 分) :

1: (10 分) 给定如下要进行聚类的元组, {3,5,11,13,4,21,31,12,26}, 要求的簇的数量为 $k=2$, 初始簇的质心为 $m_1=3$, $m_2=5$, 给出使用 K-means 算法进行聚类的步骤。

2: (10 分) 下表是由某公司雇员数据库的训练数据组成的, 数据已经进行了泛化处理。如: 年龄“31...35”表示 31 到 35 岁之间。对于给定的行, count 表示 department, age, salary 和 status 在该行上具有给定值的元组数。Status 表示分类属性。

Department	Age	Salary	Status	Count
Sales	31..35	46k..50k	senior	30
Sales	26..30	26k..30k	junior	40
Sales	31..35	31k..35k	junior	40

Systems	21..25	46k..50k	junior	20
Systems	31..35	66k..70k	senior	5
Systems	26..30	46k..50k	junior	3
Systems	41..45	66k..70k	senior	3
Marketing	36..40	46k..50k	senior	10
Marketing	31..35	41k..45k	junior	4
Secretary	46..50	36k..40k	senior	4
Secretary	26..30	26k..30k	junior	6

那么, 对于给定的一个数据元组, 它在属性 department, age 和 salary 上的数值分别为“systems”, “26..30”和“46..50k”。该元组 status 的朴素贝叶斯分类结果是什么? 给出计算过程。

三 综合题（50 分）：

1：（25 分）针对模式集{announce， annual， annually}和文本串”HIT_annual_conference_announce”，分别回答如下问题：

- 1) 对于 AC 算法，请给出转移函数和失效函数。（8 分）
- 2) 对于 WM 算法，请给出 SHIFT 表、HASH 表、和 PREFIX 表？给出文本串的匹配过程？（9 分）
- 3) 请比较 AC 算法和 WM 算法各自的优缺点？（8 分）

2（25 分）2020 年，突如其来的新冠肺炎疫情在全球暴发。由于疫情来势凶猛，在其初现之时，人们对其认知水平有限，加之在信息披露上的不及时、不透明，所以随着疫情的发展，各种传闻、流言、小道消息等不实信息蜂拥而起，四下蔓延，不仅误导了公众，还扰乱了抗击疫情的正常秩序。从传播方式和传播渠道看，此次疫情期间不实信息的传播，包括口头传播、文字传播、图片传播、视频传播及融媒体传播等多种方式，呈现出多元化局面。请从信息内容安全监管的角度回答如下问题。

- 1)：在传统的 web1.0 模式下，可通过何种技术性手段来管控互联网网站上不实信息的传播？（7 分）
- 2)：在类似微博等社交网络媒体下，有何种技术手段管控不实信息的传播。（7 分）
- 3)：在基于目录服务器和基于 DHT 结构的 P2P 网络中，有何种技术手段对不实信息进行管控，可从发布阶段、搜索阶段和下载阶段分别论述。（7 分）
- 4)：应对网络不实信息问题，请分别从管理者，运营商，服务提供者和普通网民的角度论述各种网络主体应担负的责任。（4 分）