

Задание по направлению «Информационные технологии»

Профили:

«Компьютерное моделирование в космической технике и технологиях»

«Информационные системы и компьютерные сети»

Время выполнения задания – 180 мин.

I. ОБЩАЯ ЧАСТЬ

Решите следующие задачи.

Задача 1 – 25 баллов.

По результатам анализа некоторого набора текстов получена следующая статистика встречаемости букв в русском языке.

А - 1407128	Л - 723099	Ц - 91519
Б - 273424	М - 610454	Ч - 264749
В - 788173	Н - 1219282	Ш - 100843
Г - 272734	О - 2017145	Щ - 78603
Д - 522940	П - 519662	Ъ - 7327
Е - 1502016	Р - 897631	Ы - 366437
Ж - 165154	С - 977699	Ь - 293178
З - 298896	Т - 1275553	Э - 80464
И - 1385622	У - 425206	Ю - 128557
Й - 223617	Ф - 66797	Я - 348088
К - 627560	Х - 181655	

Зная статистику распределения, расшифруйте следующую фразу.

**RQOLSQRLH OWITWQAAE SBFHSUDHRRRI OIDEPQHU DQPL PQRSE RQ
BSOHPRIH OWI-IKJHRLH ICCLAOLQJE**

Постарайтесь написать алгоритм или программу расшифровки.

Решение

Расшифрованная фраза

**НАПИСАНИЕ ПРОГРАММЫ СУЩЕСТВЕННО ПОВЫШАЕТ ВАШИ
ШАНСЫ НА УСПЕШНОЕ ПРОХОЖДЕНИЕ ОЛИМПИАДЫ**

В начале задания приведена статистика для кириллического алфавита, не содержащего диакритические знаки. В связи с этим логично предположить, что фраза изначально записана на русском языке, после чего проведена замена на буквы латинского алфавита.

При расшифровке программы может использоваться один из следующих алгоритмов.

1. Метод грубой силы. Последовательным перебором всех возможных подстановок букв пытаемся найти фразу, все слова которой находятся в словаре.

2. Здесь мы можем использовать особенности русского языка. Фраза содержит одно слово из двух букв и одно слово из четырех букв. Так как нам дан словарь, мы можем найти в нем все кандидаты на двухбуквенное и четырехбуквенное слова такие, что вторые буквы в них совпадают. Это даст нам 5 возможных букв.

Далее можно использовать информацию о статистике распределения букв в тексте. Однако так как фраза достаточно короткая, можно ожидать нарушения статистики распределения частот употребления букв в ней.

Разобьем все буквы русского алфавита на 4 группы по частоте встречаемости (число букв может отличаться, однако делать очень крупное дробление нет смысла, так как это ничего не даст, а мелкое дробление приведет к большим ошибкам). Рассчитаем частоту встречаемости латинских букв в сообщении. Это даст нам соответствие меньшего количества русских букв меньшему числу латинских. Для начала будем считать, что каждая буква попадает только в свою группу частотности. Далее можно использовать метод грубой силы с учетом соответствия групп друг другу.

Заметим, что в данной задаче соответствие соблюдается не везде, в связи с чем необходимо использовать буквы не только из данной группы, но и из соседних.

Подбор слов ведем последовательно увеличивая длину рассматриваемых слов (вместо перебора их по порядку). На каждом следующем слове используем информацию о соответствии букв друг другу, полученном для предыдущих слов. Также заметим, что постоянно приходится хранить несколько наборов соответствий и полученные на них списки слов-кандидатов. При проверке гипотез удобно отсеивать слова из этих списков.

Задача 2 – 25 баллов.

Дано: Медная витая пара, длиной 768 метров. Используемая технология ADSL2 AnnexA. В частотном диапазоне от 34,5 КГц до 43 КГц, от 138 КГц до 172,5 КГц передача данных невозможна из-за сильных постоянных помех. В частотном диапазоне от 621КГц до 966КГц есть плавающие помехи, которые снижают среднее значение бит/бод до 8 в этом диапазоне.

Необходимо выяснить, возможно ли одновременно от провайдера к клиенту передать видеопоток 5,5Мбит/с, голос VoIP 128 Кбит/с, и данные 4 Мбит/с, а от клиента к провайдеру передать голос 128Кбит/с и данные 1,2Мбит/с.

Дополнительная информация:

Частотный диапазон Upstream (от клиента к провайдеру) — от 26 до 134 КГц

Частотный диапазон Downstream (от провайдера к клиенту) — от 138 до 1104 КГц

Используемая технология — DMT (разделение всей несущей на отдельные подканалы шириной 4,3125КГц и внутри каждого подканала используется QAM)

Периодичность отправки данных в каждом подканале DMT – 4КГц.

Среднее количество бит/бод = 14 (в соответствии со стандартом и практическими применениями).

Решение

1. Расстояние 768 метров позволяет использовать ADSL2 без дополнительных ограничений.

2. Upstream $\geq \{ ABS[(134-26)/4,3125] - ABS[(43-34,5)/4,3125] - 2 \} * 14 * 4000$ (бит/с) = (25-2-2) * 14 * 4000 = 1,12 Мбит/с

Upstream $\leq \{ ABS[(134-26)/4,3125] - ABS[(43-34,5)/4,3125] \} * 14 * 4000$ (бит/с) = (25-2) * 14 * 4000 = 1,22 Мбит/с

Необходимая пропускная способность канала $(1,2*1024+128)/1024 = 1,32$ Мбит/с

Канала Upstream недостаточно для одновременной передачи данных по

условию задачи.

3. $\text{Downstream} \geq \{ \text{ABS}[(1104-172,5)/4,3125] * 14 * 4000 \} - \{ \text{ABS} [(966-621)/4,3125 + 2] * 6 * 4000 \} = 216 * 14 * 4000 - 82 * 6 * 4000 = 9,65 \text{ Мбит/с}$
Необходимая пропускная способность канала $(5,5 + 4 + 128/1024) = 9,625$ Мбит/с
Канала Downstream достаточно по условиям задачи.

II. СПЕЦИАЛЬНАЯ ЧАСТЬ

В соответствии со своим выбором программы магистерской подготовки выберите и выполните только один из следующих блоков заданий специальной части.

Блок 1. «Информационные системы и компьютерные сети»

Задача 1 – 15 баллов.

В локальной сети малого предприятия (один сетевой ввод, не более 50 ПК, выделенный UNIX-сервер для управления сетью и сервер баз данных) необходимо обеспечить следующую функциональность:

- доступ в Интернет рабочих станций;
- обмен почтой;
- управление рабочими станциями;
- ограничение доступа к серверу баз данных со стороны внешних пользователей и со стороны некоторых рабочих станций;
- файловый обмен между рабочими станциями и сервером;
- размещение на сервере учебных материалов.

Выберите и обоснуйте (в свободной форме) состав служб, устанавливаемых на выделенном UNIX-сервере.

Решение

Для обеспечения доступа в Интернет рабочих станций, которым выделяются немаршрутируемые IP-адреса, на выделенном UNIX-сервере необходимо установить службу трансляции сетевых адресов NAT (network address translation).

Для передачи сообщений от одного компьютера к другому применяются агенты пересылки сообщений (MTA, mail transfer agent). Обычно их называют «почтовый сервер», «сервер электронной почты» или «мейл-сервер». Одним из применяемых агентов пересылки сообщений для UNIX является программа Postfix, распространяемая на условиях свободного программного обеспечения. Её необходимо установить на выделенном UNIX-сервере.

В качестве почтового клиента на рабочих станциях могут использоваться, например, Outlook Express от Microsoft или Thunderbird от Mozilla.

В простейшем случае для управления рабочими станциями используется служба DHCP (Dynamic Host Configuration Protocol — протокол динамической настройки узла) (выделение рабочим станциям IP-адресов, сообщение рабочим станциям параметров, необходимых для работы в сети TCP/IP).

Для ограничения доступа к серверу баз данных со стороны внешних пользователей и со стороны некоторых рабочих станций могут быть применены:

- Аппарат идентификаторов пользователей и их паролей.
- Межсетевой экран, который для среды UNIX можно построить на базе утилиты ipfw (OS FreeBSD) или утилиты iptables (для Linux).

Наилучшим решением в случае гетерогенной локальной сети (выделенный UNIX-сервер и рабочие станции (ОС Windows)) является установка на сервере пакета Samba.

Для размещения на сервере учебных материалов рекомендуется использовать системы управления обучением (LMS – learning management systems). Наиболее популярной и развитой на сегодняшний день является LMS Moodle.

Задача 2 – 10 баллов.

Почему существуют разновидности технологий FastEthernet и GigabitEthernet с передачей по медным жилам на расстояния до 100 метров, а для технологии 10GE такой разновидности нет, при этом на оптических линиях для всех трех технологий можно использовать один и тот же кабель с разными трансиверами?

Решение

Частота пропускания по медной жиле категории Cat5E, Cat6 составляет 125 МГц по одной паре – нельзя пропустить 10GE физически из-за отсутствия кодирования уровня 20 бит/бод для расстояний 100 метров, принятых IEEE 802.3 для медных линий.

У оптической жилы физическое ограничение частоты передачи составляет 10^{14} Гц с кодированием 1 бит/бод.

Задача 3 – 25 баллов.

Решите задачу.

Дано: Два устройства, расположенные на расстоянии 45 км друг от друга. Каждое устройство имеет стабильный симметричный канал со спутником. Пропускная способность канала от Устройства 1 = 1024 Кбит/с, а от Устройства 2 = 2048 Кбит/с.

Необходимо оценить время передачи информации, объемом 15 МБайт от Устройства 2 к Устройству 1 при использовании UDP транспорта.

Справочная информация:

Технология канального уровня Ethernet со средним размером кадра 1218 байт

Технология физического уровня вносит 8 бит служебной синхронизации на каждый кадр.

Размер заголовков канального уровня — 18 байт

Размер заголовка сетевого уровня — 24 байта

Размер заголовка транспортного уровня — TCP = 24 байта, UDP = 8 байт

Средний размер пакета IP – 53465 байт

Размер служебной информации на уровнях, выше транспортного, принять равным 0.

Расстояние от любого Устройства до спутника — 35800 км

Скорость распространения сигнала от Устройств до спутника и обратно принять равной скорости света ($3 \cdot 10^8$ м/с)

Оба Устройства подключены к одному спутнику, метод передачи у спутника — Store_and_forward, скорость внутренней шины на спутнике 100Мбит/с с загруженностью менее 30%.

Решение

UDP транспорт – отсутствие подтверждения принятия сообщения удаленной стороной. Т.е. надо рассмотреть только один поток от Устройства 2 к Устройству 1.

По условию задачи не задано время «молчания» между двумя последовательно идущими кадрами, примем это время равное 0.

Необходимо определить объем реальных данных на уровне L2 с учетом служебной информации:

- на L3 количество пакетов $(15 \cdot 1024 \cdot 1024 / (53465 - 8)) = 295$, объем данных $295 \cdot 53465 = 15772175$ байт

- на L2 количество кадров $(15772175 / (1218 - 1 - 18 - 24)) = 15772175 / 1175 = 13424$.
Объем данных = $13424 \cdot 1218 = 16350432$ байт, или 127737,75 Кбит

Суммарная пропускная способность канала считается по наиболее узкому месту. В нашем случае = 1024 Кбит/с.

Время на передачу составит $127737,75 / 1024 = 124,74$ сек.

Время коммутации на спутнике можно принять = 0.

Время от земли до спутника = $35800 \cdot 10^3 / 3 \cdot 10^8 = 0,12$ сек. Дополнительная задержка = $0,12 \cdot 4 = 0,48$ сек.

Итоговое время 125,26 сек.

Блок 2. «Информационные технологии в создании аппаратуры космического базирования»

Задача 1 – 20 баллов.

Даны два объекта произвольной формы. При соударении объекты упруго отталкиваются друг от друга. Предложите алгоритм аппроксимации объектов и расчета отталкивающей силы, возникающей при соударении.

Решение

Так как для объектов произвольной формы расчет моментов сил и инерции связан с интегрированием по форме объекта (которая чаще всего бывает «неудобной») можно аппроксимировать объект набором сфер. Момент инерции при этом будет рассчитываться относительно центра масс объекта, который находится как среднее центров всех сфер.

Аппроксимация объектов сферами может проводиться, например, методом Монте-Карло, когда мы случайным образом разыгрываем центр сферы и ее радиус, после чего проверяем находится ли она внутри объекта больше, чем на заданную величину.

Задача 2 – 20 баллов.

Дан произвольный граф, дуги которого помечены буквами алфавита. Дан словарь, содержащий в себе слова, составленные из букв того же алфавита. Путь в графе задает некоторую строку. Напишите алгоритм, который позволяет найти вершину графа, служащую началом для максимального количества слов из словаря.

Решение

В данной задаче следовало для каждой вершины найти все такие пути, что конкатенация букв, лежащих на дугах, дает нам слово из словаря. Результатом будет вершина, для которой найдено максимальное количество путей.

Для определения путей используем поиск в глубину, реализуемый, например, при помощи рекурсивного алгоритма. Пусть дан путь, по которому мы пришли в данную вершину. Если данный путь является словом, прибавляем данное слово к множеству найденных слов (все слова в этом множестве должны быть уникальны).

Путь представляет собой строку. Для каждой дуги, выходящей из данной вершины, присоединяем к имеющейся строке букву, стоящую на данной дуге. Далее находим список слов, хранимых в словаре и обладающих таким префиксом. Если список не пуст, рекурсивно переходим к вершине, в которую ведет данная дуга.

Алгоритм сойдется, так как длина слова в словаре ограничена.

Мощность найденного множества слов и будет искомым числом для данной вершины.

Задача 3 – 10 баллов.

Что такое Байесовская формула полной вероятности? В каких случаях она применяется?

Решение

Ответ приведен, например, в Википедии

(см. http://ru.wikipedia.org/wiki/Теорема_Байеса).