

Школа лингвистики, 2023-24 уч. год**Дискретная математика для лингвистов****Вроде бы последний листок (Языки, автоматы и кодирование)***В. В. Кочергин, А. В. Михайлович*

Задача 1. Выяснить, является ли следующий код с кодирующим алфавитом $\{a, b, c\}$ однозначно декодируемым? Если нет, то привести слово минимальной длины, декодируемое неоднозначно.

- (a) $\{ab, ac, aaa, aabbc, b, ca, cb, ccc, ccbba\}$;
- (b) $\{ba, ab, bc, abc, cbaa, bcabb, bcaba\}$;
- (c) $\{ba, ab, abc, bac, abca, ccabba, bccc\}$;
- (d) $\{aab, acb, bac, cab, aabcb, ababcbab\}$.

Задача 2. Являются ли следующие языки над алфавитом $\{a, b\}$ регулярными?

- (a) Язык, содержащие все слова любой длины, в которых на нечётных местах находится буква b , и не содержащий других слов.
- (b) Язык, содержащие все слова любой длины, в которых на местах, номерами которых являются простые числа, находится буква b , и не содержащий других слов.
- (c) Язык, являющимся дополнением к предыдущему пункту.
- (d) Язык, содержащий все слова, кроме слов, содержащих ровно три буквы a , и не содержащий никаких других слов.
- (e) Язык, полученный конкатенацией языков из первых двух пунктов.
- (f) Язык, в словах которого одинаковое число букв a и b .

Задача 3. Ниже приведена таблица с приблизительными частотами букв русского языка в текстах.

пробел	0.175	Р	0.04	Я	0.018	Х	0.009
О	0.09	В	0.038	Ы	0.016	Ж	0.007
Е, Ё	0.072	Л	0.035	З	0.016	Ю	0.006
А	0.062	К	0.028	Ь, Ъ	0.014	Ш	0.006
И	0.062	М	0.026	Б	0.014	Ц	0.003
Т	0.053	Д	0.025	Г	0.013	Щ	0.003
Н	0.053	П	0.023	Ч	0.012	Э	0.003
С	0.045	У	0.021	Й	0.010	Ф	0.002

Построить оптимальный код для заданных частот в случае, когда кодирующий алфавит состоит из пяти букв.