

0. Выбрать большой литературный текст (или несколько текстов одного автора) в максимально простом формате (.txt). Объём 0.5 - 2 Мб. Лучше брать на английском (или любом другом языке, все символы которого есть в ASCII). Прислать мне сам текст или ссылку, тем самым застолбив этот текст за собой.

1. Сделать декапитализацию: убрать все большие буквы, закодировав их позиции максимально компактно (указать способ кодирования с обоснованием его выбора и длину кода). Указание: возьмите набор правил капитализации, постройте битовый массив, указывающий позиции всех отклонений от правил и закодируйте его как набор чисел. (Возможные правила капитализации: первая буква текста; буква, следующая за точкой, с пробелами/переводами строки или без; буква, следующая сразу за двумя заглавными. Можно указать свои правила.)

2. Для декапитализированного текста построить все контекстные модели 3-го порядка (для каждого контекста собрать всю информацию, как в лекции 7). Посчитать необходимый объём памяти для хранения всех моделей.

Оба задания прислать с описанием не только результата, но и хода решения.