Поразмыслил я и над тем, как при заданном распределении вероятности длины кодовых слов построить примитивный префиксный код (или соответствующее ему дерево) с переменной длиной слова, чтобы средняя длина кодового слова была наименьшей. При $N=2$ вопрос тривиален. Если $N=3$, то единственному примитивному префиксному коду, принимаемому в расчет, соответствует дерево, изображенное на рис. 4. Остается лишь выяснить, какую из вероятностей p_1, p_2, p_3 следует поставить в соответствие вершине графа, находящейся на расстоянии, равном 1, от корня дерева. Нетрудно видеть, что ею должна быть наибольшая вероятность, то есть двум вершинам графа, отстоящим от корня на расстояние, равное 2, мы должны поставить в соответствие две

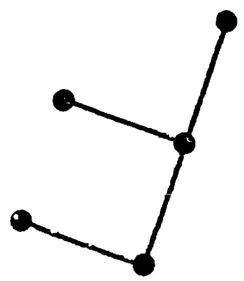

РИС. 4.

меньшие вероятности. Это и навело меня на мысль о том, как в общем случае можно построить код с минимальной средней длиной кодового слова или соответствующее ему дерево. Построение я производил рекуррентным способом. Предположим, что при числе кодовых слов, равном N, способ построения уже известен. Зададим теперь $(N+1)$ кодовых слов и расположим их в порядке убывания их вероятностей $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_{N-1} \geqslant p_N \geqslant p_{N+1}$. Прежде всего заменим исходное распределение вероятности другим. Для этого вычеркнем две наименьшие вероятности p_N и p_{N+1} и вместо них припишем к числам $p_1, p_2, \ldots, p_{N-1}$ новое число $p_N^* = p_N + p_{N+1}$. Проделав это преобразование, известное под названием «сжатие алфавита», мы получим распределение вероятности, состоящее из N членов. По предположению способ построения примитивного префиксного кода с минимальной средней длиной кодового слова или соответствующего такому коду дерева с N свободными вершинами, которым соответствуют числа $p_1, p_2, \ldots, p_N^*$, уже известен. В этом дереве из вершины, помеченной числом p_N^*, выпустим две ветви, а их концам поставим в соответствие числа p_N и p_{N+1}. Мы получим префиксный код с минимальной средней длиной кодового слова для распределения вероятности слов $(p_1, p_2, \ldots, p_N, p_{N+1})$.

Поясним построение кода на примере. Пусть $N=5$, а распределение вероятности слов имеет следующий вид:

$$p_1 = \frac{1}{3}, \quad p_2 = \frac{1}{5}, \quad p_3 = \frac{1}{5}, \quad p_4 = \frac{1}{6}, \quad p_5 = \frac{1}{10}.$$

Вычислив сумму двух меньших вероятностей, получим распределение $^1/_3$, $^4/_{15}$, $^1/_5$, $^1/_5$. Вычислив сумму двух меньших вероятностей в новом распределении, получим трехчленное распределение $^2/_5$, $^1/_3$, $^4/_{15}$. Что делать с распределением, состоящим из трех чисел, известно. Так мы получаем дерево кода, которое представлено на рис. 5, с средней длиной слова $^{34}/_{15} = 2{,}266\ldots$ знаков (заметим, что $\sum_{k=1}^{5} p_k \log_2 (1/p_k) = 2{,}220\ldots$) (На рис. 5 число, стоящее у любой точки ветвления, равно сумме чисел, стоящих у висячих (свободных) вершин, до которых можно добраться, поднимаясь из данной точки ветвления вверх по дереву.)

Я продумал до мельчайших подробностей доказательство для общего случая. Это оказалось совсем не трудно. Не стану приводить его здесь; уже после того, как мне удалось решить задачу и найти доказательство для общего случая (чему я был очень рад), в одной из книг, которую дал мне почитать лектор, я обнаружил «свой» метод построения кода и узнал, что префиксный код с минимальной средней длиной кодового слова и заданным распределением вероятности слов называется *кодом Хафмана.*

Не скрою, узнав, что построенный мной префиксный код с минимальной средней длиной слова давно известен, я был несколько разочарован. Мне так хотелось открыть нечто новое! Однако, поразмыслив над случившимся, я понял, что ничего другого ожидать было нельзя: ведь мы едва успели познакомиться с азами теории информации, и рассчитывать на то, что на этом уровне нам встретится нерешенная задача, было по меньшей мере наивно. Впрочем, я не сожалею о времени, затраченном на решение задачи, поскольку успел уяснить для себя многое из того, что касается кодов и их деревьев, и это позволило мне глубже разобраться в вопросах, затронутых на лекции. Где-то я вычитал, что человек до конца понимает лишь то, до чего додумывается сам, подобно тому как растение усваивает лишь ту влагу, которую впитывают его корни.

Сегодня же мне довелось побеседовать с нашим лектором и рассказать ему о том, как я «открыл» код

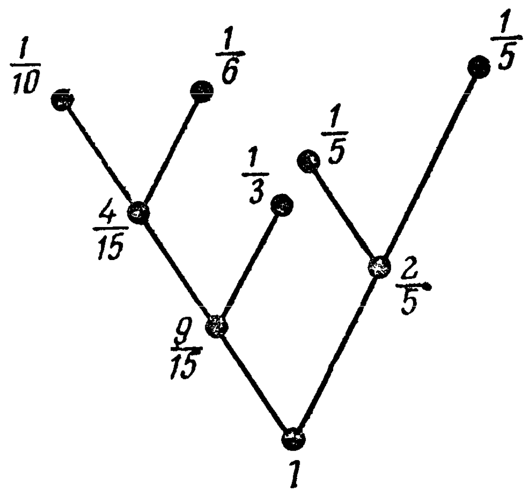

РИС. 5.

Хафмана. В утешение мне лектор заметил, что, как бы там ни было, мне удалось получить замечательный результат и то, что я не был первым, отнюдь не умаляет значимости найденного мной решения. В связи с кодом Хафмана лектор упомянул об одной действительно не решенной проблеме: как следовало бы видоизменить построение кода с минимальной средней длиной слова, если длина любого кодового слова не превосходит заранее установленного предела? Я непременно подумаю над этим вопросом, хотя трудно сказать заранее, почему он столь труден. Ясно лишь, что предложенная лектором задача не из легких, иначе ее бы давно решили. Эта задача показывает, что довод, приведенный мной выше, не верен: хотя мы успели познакомиться лишь с простейшими понятиями теории информации, но уже встретили задачу, решение которой пока не известно!