Префиксные коды для кодирования целых чисел [1]

1. Описание множеств кодовых слов

[. . .]

Приведем краткое описание побуквенных кодов, которые были исследованы экспериментально. Предварительно введем следующие обозначения: $\beta(m)$ – запись целого числа m, $m \geq 0$, в двоичной системе счисления (начиная с первой левой единицы); $\hat{\beta}(m)$ – $\beta(m)$ без первой левой единицы; $L(x)$ – длина слова x; $\{0\}^n$ – строка, состоящая из n нулей; $\alpha(m) = \{0\}^{L(\beta(m))-1}$.

Например, $\beta(9) = 1001$; $\hat{\beta}(9) = 001$; $L(\beta(9)) = 4$; $\{0\}^2 = 00$; $\alpha(9) = 000$.

Тогда простейший исследованный код, состоящий из бесконечного множества кодовых слов и обозначаемый как $\varphi_0(m) = \{0\}^m 1$, где $m \geq 0$ – целое. Например, $\varphi_0(3) = 0001$.

Код φ_0 является основой других, более сложных исследуемых кодов.

Так, следующее множество кодовых слов, которое мы обозначим через φ_1, использует множество φ_0 и задается последовательностью двух подслов: $\varphi_1(m) = \alpha(m+1)\beta(m+1)$. Например, $\varphi_1(9) = 0001010$.

Используя для кодирования подслова $\alpha(m+1)$ код φ_1, можно получить другое кодовое множество **[1]** (обозначим его через φ_2), слова которого представляют собой последовательность трех подслов: $\varphi_2(m) = \alpha(k)\beta(k)\hat{\beta}(m+1)$, где $k = L(\beta(m+1))$. Например, $\varphi_2(9) = 00100010$.

Кодируя подслово $\alpha(k)$ кодом φ_1 можно получить из кода φ_2 новое кодовое множество φ_3 **[1]**. Кодовые слова множества φ_3 задаются последовательностью четырех подслов: $\varphi_3(m) = \alpha(p)\beta(p)\hat{\beta}(k)\hat{\beta}(m+1)$, где $p = L(\beta(k))$. Например, $\varphi_3(9) = 01100010$.

Таким образом, каждый раз кодируя число нулей α кодом φ_1, можно получать все новые и новые коды. Например, $\varphi_4(9) = 010100010$.

Таблица 1

Основные кодовые множества для натуральных чисел

Код	Определение кодового слова для числа m ($m \geq 0$)	пример кодового слова для $m = 4$
φ_0	$\{0\}^m 1$	00001
φ_1	$\alpha(m+1)\beta(m+1)$	00 101
φ_2	$\alpha(k)\beta(k)\hat{\beta}(m+1)$, где $k = L(\beta(m+1))$	0 11 01
φ_3	$\alpha(p)\beta(p)\hat{\beta}(k)\hat{\beta}(m+1)$, где $p = L(\beta(k))$	0 10 1 01
φ_4	$\alpha(q)\beta(q)\hat{\beta}(p)\hat{\beta}(k)\hat{\beta}(m+1)$, где $q = L(\beta(p))$	0 10 0 1 01

В табл. 1 приведены определения и примеры кодовых слов из множеств φ_0, φ_1, φ_2, φ_3, φ_4 для натуральных чисел m.

Для описания кодовых множеств, предназначенных для кодирования конечных алфавитов, введем дополнительные обозначения: $N(\varphi)$ (или N) – количество слов в кодовом множестве φ; $\gamma(m,t)$ – двоичная запись натурального числа m, в которой берется t крайних справа знаков. Если $\beta(m) < t$, то $\beta(m)$ дополняется слева нулями. Например, $\gamma(5,2) = 01$, $\gamma(5,8) = 00000101$.

[1] Из статьи: *Е. И. Ситняковская*, Построение эффективных побуквенных кодов для словарных методов сжатия данных. // Проблемы передачи информации, том. 34, вып.2, Апрель-Май-Июнь 1998, С. 47-55.

Тогда одно из конечных кодовых множеств (обозначим его через φ_5), впервые предложенное в **[2]**, можно задать следующим образом:

$$\varphi_5(x) = \gamma(k,t)\hat{\beta}(m+1), \tag{1}$$

где $k = L(\hat{\beta}(m+1))$, $t = \lceil \log\log N \rceil$ при нецелом $loglogN$, и

$$\varphi_5(m) = \begin{cases} \gamma(k,t)\hat{\beta}(m+1) & \text{при } m = 0,\dots,N-3, \\ 1^t\hat{\beta}(N-1)0 & \text{при } m = N-2, \\ 1^t\hat{\beta}(N-1)1 & \text{при } m = N-1, \end{cases} \tag{2}$$

где $k = L(\hat{\beta}(m+1))$, $t = \log\log N$ при целом $\log\log N$.

Например, при $N(\varphi) = 16$ значение $\log\log N$ – целое. Следовательно, для построения кода φ_5 используем формулу (2), при этом $t = \log\log 16 = 2$ ($\varphi_5(0) = 00$, $\varphi_5(1) = 010$, $\varphi_5(2) = 011, \dots, \varphi_5(15) = 111111$).

Удобным представлением множества кодовых слов являются двоичные деревья. На рис. 3 изображено дерево, соответствующее коду φ_5, количество N слов которого равно шестнадцати.

На базе кода φ_5 автором построены новые конечные кодовые множества. Ниже приведем их описание.

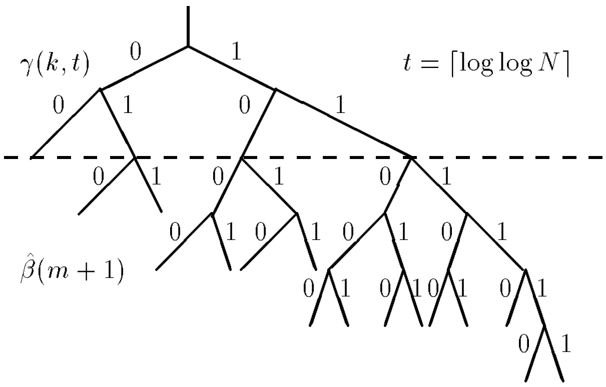

Рис. 3. Представление в виде дерева кода φ_5, состоящего из шестнадцати слов ($N = 16$)

Добавив к каждому концевому узлу дерева (см. рис. 3) по два сына или, другими словами, "расщепив" эти узлы на два, получим новое кодовое множество (обозначим его через φ_5^1), состоящее из тридцати двух слов. Таким образом, можно сказать, что код φ_5^1 получен из кода φ_5 дляалфавита объема $N = 16$ путем одного расщепления.

Расщепляя кодовое множество φ_5, состоящее из шестнадцати слов, два, три и более раз, можно получать, соответсвенно, множества φ_5^2, φ_5^3, φ_5^4 и т.д.

Опишем эти множества более формально. Для этого обозначим количество расщеплений через n. Тогда код φ_5^n для целого m, $0 \le m \le N-1$, задается равенством $\varphi_5^n(m) = \varphi_5(\lfloor m/2^n \rfloor)\gamma(m - 2^n(\lfloor m/2^n \rfloor), n)$. Например, при $N = 32$ получаем $\varphi_5^1(5) = 0111$.

На основе расмотренного выше кода φ_5 и его модификаций автором было разработано еще одно кодовое множество (обозначим его через φ_6), предназначенное для кодирования конечных алфавитов. Код φ_6 может быть задан следующим образом:

$$\varphi_6(m)=\begin{cases}\gamma((L(\hat\beta(k+m))-\log k),t)\hat\beta(k+m) & \text{при } m=0,\dots,(N-2)-k,\\ \{1\}^t\hat\beta(N-1)\gamma(m-N+1+k,\log k) & \text{при } m=(N-1)-k,\dots,N-3,\\ \{1\}^t\hat\beta(N-1)\beta(k-1)0 & \text{при } m=N-2,\\ \{1\}^t\hat\beta(N-1)\beta(k-1)1 & \text{при } m=N-1,\end{cases}$$

где $t=\lceil\log\log N\rceil$ при нецелом $\log\log N$, $t=\log\log N-1$ при целом $\log\log N$, $r=2^t$, $k=\frac{N}{2^r}$. Например, при $N=8$ получаем $\varphi_6(3)=101$.

Кроме неравномерных кодовых множеств исследовался также и равномерный код (обозначим его через φ_7), который можно задать следующим равенством:

$$\varphi_7(m)=\gamma(m,t),\quad \text{где } t=\lceil\log N\rceil.$$

Например, при $N=8$ имеем $\varphi_7(5)=\gamma(5,3)=101$; при $N=15$ имеем $\varphi_7(5)=\gamma(5,4)=0101$.

Список литературы

[1] *Bell T. C., Witten I. H., Cleary J. G.* Text compression. Englewood Cliffs, N.J.: Prentice-Hall, 1990.

[2] *Рябко Б. Я.* Дважды универсальное кодирование // Пробл. передачи информ. 1984. Т.20 №3. С.24-28.