Scan A7178

DE Knuth
letter to RE Tarjan
and
NJA Sloane

2 pages, 1 seq

Asked SP to do

to N.J.A. Sloane
(I found this in old files and thought you might be interested...)

S94

A7178

July 22, 1975

Dr. R. E. Tarjan
c/o Don Rose
Aiken Computation Laboratory
Harvard University
Cambridge, Mass. 02138

Dear Bob:

As I told you in June, I made a little progress on the Huffman-coding lower bound. Let $F(n)$ be the number of essentially distinct binary prefix codes, where two codes are "essentially distinct" if they assign a different length codeword to some message. Thus

n =	1	2	3	4	5	6
$F(n)$ =	1	1	3	13	75	525

and $F(n)$ is the sum of $n!/\ell_0!\,\ell_1!\,\ldots$ over all nonnegative integer vectors $(\ell_0,\ell_1,\ldots)$ where $\ell_0+\ell_1+\ldots = n$ and $\ell_0 + \frac{1}{2}\ell_1 + \frac{1}{4}\ell_2 + \ldots = 1$.

It follows that $F(n)$ is the coefficient of z^{2^n} in the polynomial $(z^1+z^2+z^4+\ldots+z^{2^{n-1}})^n$. From this representation it is obvious that $F(n) \leq n^n$. Hence the best conceivable information-theoretic lower bound will be of the form $n \lg n = \lg n! + n \lg e$; we're getting $1.44n$ more than the lower bound for sorting, while your upper bound is the sorting time plus about $2n$.

I believe I can get an asymptotic formula for $F(n)$ using complex analysis, and it will probably be something like $F(n) \sim c\,n^{n-k}$ for constants c and k. Thus the information-theoretic bound will indeed be of the form $n \lg n + O(\lg n)$.

In other words, an improvement on Huffmann's procedure which does, say, n operations of $\lg n$ binary decisions each isn't out of the question; but if my analysis works out as expected, an improvement of the form $S(n) + n$ will be impossible, even on the average, since $S(n) \leq n \lg n - 1.329n + O(\log n)$ [exercise 5.3.1-15]. Of course the average time can be reduced if we choose our distribution of essentially-distinct codes to be sufficiently nonuniform.

Cordially,

Donald E. Knuth
Professor

P.S. Can the exact value of $F(n)$ be computed in polynomial time (a polynomial in n not $\log n$)?

P.P.S. The values of $F(n)$ for $n \leq 5$ agree with those of P_{n-1} in my exercise 5.3.1-4, so I thought for a minute that a surprising result was going to turn up. But $n = 6$ shot this down.

DEK/pw