A 32-base numeral system

With the ongoing expansion of technology, particularly in information and communication, there is increasing use of very large databases. Each member of a very large set is often identified by a number in Arabic/Hindu numeral system or an alphanumeric code. The identifier is unique in the sense that there is supposed to be at most one object for each identifier, and one and only one identifier for each object. The function of the identifier is to enable dealing with information on the set of objects without ambiguity and with efficiency. Postal codes, vehicle registration numbers, student roll numbers in national examinations, identification numbers of library books, passports, currency notes and share certificates are but a few examples. When the number of members of the set increases beyond several million, the number of digits or characters, that have to be used, becomes sufficiently large for user-friendliness to become a serious concern. Larger the number of digits or symbols used for identification, less user-friendly is the system. An extreme case to illustrate the point is that of a very large Indian financial institution using numbers of 15 digits for identification of its certificates! Fortunately, ordinary people are not often required to use such formidable numbers.

While codes based on alphanumeric characters have advantages of compactness, the efficiency of processing them is rather limited. Hexadecimal numbers used for identifying memory locations in a computer, unlike the above two types of choices, offer the advantages of compactness as well as efficiency in information processing. In the same spirit, a 32-base system is proposed here which seems to have advantages in future applications such as developing national databases for countries with populations of the order of a billion (e.g. for census) and, perhaps in the not-too-distant future, global databases. A number in the 32-base system would generally need two digits, where a 10-base system of numerals, that we ordinarily use, would need three. If one considers a set of roughly a billion members, one has then the choice of using a nine-digit number in the usual system as against a six-digit number in the proposed system. Clearly, if these numbers are going to be frequently used by ordinary people, the second choice would make smaller demands on the human memory and, therefore, has greater promise of widespread acceptability.

The 32-base system

The numerals, or the basic symbols representing numbers, in the proposed system consist of 10 Arabic/Hindu numerals and 22 letters of the English alphabet. They are given in Table 1 with equivalent numbers in the 10-base system.

In this scheme, four letters of the alphabet, namely, G, I, O, and S, have been excluded on the ground of their resemblance to the Arabic/Hindu numerals 6, 1, 0 and 5.

This system could be viewed as a natural extension of the hexadecimal system, each symbol in the two systems corresponding to a binary number of 5 and 4 digits respectively. This feature makes it very simple to transform a number in the present system to the equivalent binary number or to change the latter into the former. The use of increasing order of the Arabic/Hindu numerals combined with the alphabetical order of the letters in the above scheme should facilitate manual sorting and searching of vast databases.

Let $a_N \ldots a_n \ldots a_0$ denote an integer in the present system, where each digit a_n is a numeral given in the Table 1 with the value p_n. The value p of this integer in the 10-base system is then given by

$$p(a_N \ldots a_n \ldots a_0) = \sum_{n=0}^{N} 32^n p_n.$$

Similarly, a floating point number is denoted by $a_N \ldots a_n \ldots a_0.a_{-1} \ldots .a_{-M}$ and its value p is given by

$$p(a_N \ldots a_n \ldots a_0 \cdot a_{-1} \ldots a_{-M}) = \sum_{n=-M}^{N} 32^n p_n.$$

A few statements are given below to give a flavour of the new system. For the sake of clarity, the numbers in the new system are given below in bold italics so that they are not confused with those in the 10-base system, especially when the former has only Arabic/Hindu numerals.

$\boldsymbol{2A + B = Z}$.

$\boldsymbol{A^2 = 34}$.

1 KB of memory has exactly ***100*** bytes.

The compactness of the above system is illustrated by two simple examples. If one wished to represent a date in the Gregorian calendar in a period of 1024 years, starting, say, on 1 January 1501, one could use the convention ***dmyy*** or ***yymd***, where ***d***, ***m***, and ***yy*** are numerals in the present system representing the date, the month, and the year minus 1500. 15 August 1947 can then be represented as ***F8DZ*** or ***DZ8F***. The postal index number (PIN) 560 080 of the Indian Academy of Sciences, Bangalore can be written in the present system as ***J2YH***.

It is not difficult to foresee potential applications of the new system (integers as well as floating point numbers) not only in computer programing, but also in other areas like encryption, high precision computations, reporting of large exact numbers (e.g. large prime numbers) high precision approximations (e.g. values of irrational numbers) in mathematical investigations and results of high-precision measurements (e.g. in geodesy and astronomy).

Table 1. Numerals in the 32-base system and their values in the 10-base system

Numeral	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Numeral	H	J	K	L	M	N	P	Q	R	T	U	V	W	X	Y	Z
Value	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31

KIRIT S. YAJNIK

2/1 Bride Street,
Langford Town,
Bangalore 560 025, India