INTEGER SEQUENCES, ON-LINE ENCYCLOPEDIA OF

The On-Line Encyclopedia of Integer Sequences (URL: www.research.att.com/~njas/sequences/) is a large data-base of number sequences. Three examples will illustrate some of the ways in which it has been used.

(i) You discover what may be a new algorithm for checking that a file of medical records is in the correct order. To handle files of 1, 2, 3, 4, ... records, your algorithm takes 0, 1, 3, 5, 9, 11, 14, 17, 25, ... steps. However, when you look this up in the data-base, you discover that this is the number of steps needed for "sorting by list merging". The entry directs you to Section 5.3.1 of Volume 3 of D. E. Knuth, *The Art of Computer Programming*, where you find your algorithm described, as well as an explicit formula for the nth term. You decide not to apply for a patent.

(ii) You are working on one of the classically hard problems in discrete mathematics, the enumeration of Latin squares of order n. With considerable effort you find the number of order 10, which is

$$75807214831601328114892800 .$$

You consult the data-base and discover that the same number was recently computed by someone else. You end up by writing a joint paper, thanking the data-base for bringing you together (McKay and Rogoyski, 1995).

(iii) You encounter a binomial coefficient sum:

$$\sum_{k=0}^{n} \binom{4n+1}{2n-2k} \binom{n+k}{k} .$$

You know there are powerful methods for evaluating such sums by computer (Petkovšek et al., 1996), but you are in a hurry, so you work out the sum for $n = 0, 1, 2, \ldots$, obtaining

$$1, 12, 240, 5376, \ldots .$$

The data-base tells you this is $4^n \binom{3n}{n}$, and directs you to a reference that shows this is equal to your sum.

The data-base presently contains about 50,000 number sequences. A typical entry will supply a description, the first 50 or so terms, and (when available) formulae, recurrences, generating functions, references, computer code for producing the sequence, links to relevant web sites, etc. The data-base has been described as one of the most useful resources on the Internet. It is a kind of mathematical "fingerprint file" (Cipra, 1994).

The history of the data-base encapsulates the whole modern history of computers: it began in the mid-1960's on punched cards, became a book in 1973 (Sloane, 1973), was transferred to magnetic tape in 1970's and to magnetic disks in 1980's. A second book, twice the size of the first, appeared in 1995 (Sloane and Plouffe, 1995).

Until the mid-1990's the two books provided the main form of access to the data-base. However, in 1994 two electronic-mail servers were set up to provide remote access. The first (se-

quences@research.att.com) simply looks up a sequence, while the second (superseeker@research.att.com) tries hard to find an explanation even for sequences not presently in the data-base.

In 1996 the whole data-base was placed on the Internet at the address mentioned at the beginning of this article. Storage space is no longer a problem, and the data-base has grown to more than 10 times the size of the 1995 book.

Several other developments (besides improved storage facilities) have helped make the data-base possible:

- Editing. Making changes is far easier with *vi* than on a punched card machine. The programmable version of *vi*, *ex*, is extensively used in processing sequences.
- Searching. *Unix* tools such as *awk*, *grep* and the *shell* make it possible to search the data-base efficiently, both to test if a sequence is present and to retrieve a sequence.
- Arithmetic. Initially many of the sequences were produced using *Fortran*, but now the main languages used are *Maple* and *Mathematica*. Sequences are thus generated more easily and reliably.
- Modems. Initially the data-base was contained in a box of punched cards and could only be updated in a computer room. Now this can be done almost anywhere.
- Email. Formerly, new sequences would arrive in letters, often hand-written, or in preprints or reprints. Now they arrive electronically, via email or the web pages.
- Html. Sequences entered on the web page are processed via *cgi* scripts and *shell* programs. New sequences are entered on an html *form*, processed by another *cgi* script, and sent via email to the author's home computer where they are further processed. Email is used for this purpose, since the web page is on a public machine outside the *firewall* surrounding the author's home computer.

The net result is that the data-base is easily consulted: it currently receives more than 2500 hits per day, or almost a million a year, and *new* sequences continue to arrive at about 10,000 per year. Without the innovations in computers in the past 35 years this would not have been possible.

1 References

1973. Sloane, N. J. A. *A Handbook of Integer Sequences*. New York: Academic Press.

1994. Cipra, B. Mathematicians get an on-line fingerprint file, *Science* **265** (22 July), p. 473.

1995. McKay, B. D. and Rogoyski, E. Latin squares of order 10. *Electron. J. Combin.* **2**, #N3.

1995. Sloane, N. J. A. and Plouffe, S. *The Encyclopedia of Integer Sequences.* San Diego: Academic Press.

1996. Petkovšek, M., Wilf, H. S. and Zeilberger, D. $A = B$ Wellesley, MA: Peters.

N. J. A. SLOANE