Reprinted from *Journal of Range Management*
Vol. 34, No. 3, May 1981, p 239

The Binary Search for Accuracy in Plant Symbols

R.P. GIBBENS AND THOMAS N. BILAN

Abstract

The transfer of data from field sheets to computer files always involves the risk of errors being made in plant symbols or other identifying codes. If a master list of symbols or codes is incorporated into the software programs used for data entry on CRT computer terminals the risk of making errors can be substantially reduced. This is done by utilizing the highly time efficient binary search to compare each entered symbol or code with the master list. Detection of errors before the data are transferred to computer files saves much time which would otherwise be spent in retrieving and correcting files.

Interactive CRT (cathode ray tube) terminals and the ability of computers to make high-speed binary searches can be combined to reduce errors substantially when vegetation data are placed in computer files. The use of CRT terminals for data entry to computers is becoming increasingly common because this method is more time efficient than entry of data via punched cards.

Plants have long been identified by symbols or codes consisting of the first two letters of the generic and the specific scientific names. Numerals are added at the end of the codes to distinguish among duplicates in the four-character symbols. Accuracy is essential in assigning plant symbols and in transferring them to computer files.

We found that data entry errors can be greatly reduced and even some field recording errors detected by utilizing binary search procedures. The method is simple. First, a plant symbol list is compiled containing all of the plants likely to be encountered on the study area in question. This list, arranged alphabetically, is incorporated into a question-answer data entry software program. When entering field data, questions appear on the CRT in the following order: Date? Treatment? Quadrat number? Plant species? When we enter a plant symbol in answer to the plant species question, the computer checks the entered symbol, via a binary search, for a match in the master list. Even on a relatively slow minicomputer, with a master list of 320 symbols, the process takes but a fraction of a second. If a match is found, the program proceeds to the next question which deals with the attribute of interest, e.g., number, cover or weight.

If a match is not found, various error messages can be used. We have found that if the error message displayed on the CRT contains about eight symbols from the master list that alphabetically precede the entered symbol and eight symbols that alphabetically follow it, the correct symbol can often be quickly found and reentered. This is where field sheet errors can be detected. Generic name symbols are usually correct on field sheets but it is not unusual to find incorrect specific name symbols.

This method of entering data is based on the high efficiency of the binary search discussed by Knuth (1973). A binary search compares the entered symbol to the middle symbol in the master list. This probe tells which half of the list should be searched next. The same procedure compares the entered symbol to the middle symbol of the selected half, etc. After a maximum of about $\log_2 N$ (N = number of items in list being searched) comparisons, either a match for the entered symbol is found or not found. The average number of comparisons in a search is about half of $\log_2 N$.

Since it is possible to transpose one plant symbol into another legitimate symbol when entering data, there is still a need to compare computer listings of entered data with the field sheets. With the software programs we have developed, this is done twice. After data for a given quadrat or transect have been completely entered, they are displayed on the CRT in tabular form so that a quick check of entries can be made and corrections entered before the data are placed in a file. The data are also printed on paper so that another check with the field sheets can be made. The use of this type of data entry has greatly reduced the amount of time spent in retrieving and correcting files. Any lengthy list of items identified with individual codes which must be entered repeatedly, e.g., cattle identification numbers, lends itself to this method of error reduction.

Copies of a data entry program incorporating a binary search and written in BASIC for a 32K RAM minicomputer are available upon request from the authors.

Literature Cited

Knuth, Donald E. 1973. The art of computer programming Vol. 3. sorting and searching. Addison-Wesley Pub. Co. Inc. Reading, Mass. 722 p.

Authors are respectively, range scientist and data analyst, U.S. Department of Agriculture, Science and Education Administration, Agriculture Research, Jornada Experimental Range, Las Cruces, New Mew Mexico 88001.

This report is Journal article 745, Agricultural Experiment Station, New Mexico State University, Las Cruces, New Mexico 88003.

Manuscript received December 5, 1979.