WOOD IDENTIFICATION VIA COMPUTER

by

Regis B. Miller

Forest Products Laboratory*, Forest Service, P.O. Box 5130, Madison, Wisconsin 53705, U.S.A.

Summary

The various procedures used in wood identification are reviewed. A computer program (IDENT4, prepared by Larry Morse) adapted especially for wood identification is described and discussed, with reference to the work of an IAWA Committee working towards a standard list of characters suitable for computerized hardwood identification.

Introduction

Methods of wood identification have been slow to change. Earliest identification were made by experts who could recognize different woods using various judgmental techniques, such as weight, color, smoothness, luster, hardness, etc. Experience was a necessity. When dichotomous keys were introduced, experts no longer entirely depended on immediate recognition. In addition, the novice or less experienced could attempt an identification by following the correct path through a dichotomous key and then, by comparing their answer with known samples, arrive at a positive identification. Today the dichotomous key is still the most widely used method of wood identification, but other methods have been devised. For a review of these methods consult Pfeiffer and Varossieau (1945) and Varossieau (1948). The most successful method, other than the dichotomous key, was the multiple entry, marginally perforated card system proposed by Clarke (1938) and adopted by such wood anatomists as Brazier and Franklin (1961), Phillips (1941), Normand (1946), and Dadswell and Eckersley (1941). Clark’s card system was an adaptation of a keycard system earlier proposed by Bianchi (1931). For a review of the development of Clark’s cards see Brazier (1976).

In a perforated card system each card represents one taxon. Initially, the cards have holes evenly spaced along the margins and each hole represents a character or character state. If a hole is opened or slotted, using a specially designed punch or clipper, that character or character state is positive. Negative character states are represented by an unclipped hole. To identify an unknown, the state of any character is determined and then a rod is pushed through all the cards for the hole representing that character state. With the rod in the hole, all cards are lifted. Those with an open slot (positive) fall and those with a hole (negative) remain on the rod. This process continues until only one taxon remains. The biggest advantage of this system is the multiple entry capabilities, i.e., a polyclave. The operator selects the character. In a monoclave the dichotomous key dictates the character selection.

The disadvantages of the multi-entry card system are the limited number of characters that can be placed around the margins of the cards, the deterioration of the cards due to the rod, and the limited number of taxa (cards) that can be conveniently manipulated. The biggest complaint and frustration, however, occurs when cards do not fall out when they should. Also data in the punched-card system must be rearranged for published articles.

Data may be printed as a chart or anatomical data matrix (Phillips, 1941; Kukachka, 1960) or as descriptions (Brazier and Franklin, 1961). When used with a matrix, taxa are listed down the side of the chart and the characters across the top. Instead of a rod, a straight edge and a sharp eye search the matrix for the correct combination of characters. Charts or data matrices are even more limiting as an identification method than punched-card systems.

The first step toward automated wood identification systems was the proposed conversion of marginally perforated cards to 80-column computer cards which discarded the rod for a card-sorting machine. Varossieau (1948) converted the Dutch method (Pfeiffer and Varossieau, 1945) to an 80-column card known as the Hollerith System. About the same time,

* Maintained at Madison, Wis., in cooperation with the University of Wisconsin.
B.F. Kukachka at the U.S. Forest Products Laboratory developed a system using 80-column IBM cards. At the I.A.W.A. meetings in Stockholm in July 1950, Varossieu and Kukachka presented a joint paper entitled 'Suggestions for a Standardized Method of Identification with Mechanically Operated Punched Cards'. Since then, however, very little has been accomplished on such an automated punched-card system.

Today there is still no standardized method of wood identification, and the problems are the same as or worse than they were in 1950. Industrial and scientific researchers, customs officials, and manufacturers often require a proper wood identification. More and more species are being brought onto the market yearly, and many of these were uncommon in 1950. Anthropologists, archeologists, geologists, and climatologists abound, and their digging uncovers fragments of wood in various states of decay, the identification of which is important in their research. Today sophisticated forensic scientists analyze a wide variety of wood particles that may be important pieces of evidence. The antique dealers, the artists, curators of museums, hobbyists, and others who work with wood want an accurate identification of their specimens. Last year, the Center for Wood Anatomy Research at the U.S. Forest Products Laboratory identified approximately 5,000 tropical and temperate wood specimens from such a variety of clients. After years of experience, one can readily recognize some of the more common woods, whereas other woods require the use of available keys. The identification of little-known woods, small samples, or samples without a common name or origin may be difficult, if not impossible, even with keys, marginally perforated punched-cards, and many years of experience. To help alleviate this continual problem of too many specimens for identification and of lengthy or no positive identification, a system of computer programs developed for the identification of biological specimens was investigated (Morse et al., 1971; Morse, 1974).

Synopsis of Programs

In 1974 Larry Morse published a package of programs that not only contained an identification program (IDENT4), but also programs that perform the following operations: Key construction (KEY2), description printing (DSCRBE), taxon comparisons (CMPARE), data-file inversions (INVERT), and punched-card key production (CARDKEY). This system was adopted by the Seed Laboratory at the Beltsville Agriculture Research Center (BARC) in Beltsville, Md., and is now being used successfully (Gunn and Seldin, 1977). At the present time, we at the Forest Products Laboratory have converted only IDENT4 to the UNIVAC 1110. Although we have expanded the limits for the number of taxa that the program can handle, the IDENT4 we use remains basically the same as published by Morse (1974).

Data collected for the data files are coded and then aligned in the computer in an array or data matrix. These matrices are similar to the charts or tables with the taxa listed down the side and the characters across the top. Instead of a straight edge and sharp eye, it is the computer that searches the data matrix and eliminates taxa that do not match the selected character state.

Since Morse was interested in taxonomic and not anatomical data matrices, he envisioned a hierarchical system of data matrices. Thus, each taxonomic data matrix would cover a particular group of taxa, such as families in an order, genera in a family, and species in a genus. Each data matrix would then be linked to form a hierarchical system. For wood it is impractical to have separate data matrices for orders or families due to the great variability within these taxonomic groups. Data matrices of geographic origin are also not feasible since the origin of wood samples is often unknown. The construction of several smaller data matrices on the basis of characters such as scalariform perforation plates, septate fibers, and ring porosity is feasible and may be practical due to the higher cost of searching one large data matrix. On some computers it may be necessary to have several data matrices if the memory storage capacity is limited.

Types of Characters

As with any computer program, input or raw data must be organized into a set format such as a data matrix. The computer can recognize a coded character state by the position within the matrix. In Morse's system there are three types of characters: dichotomous, multistate, and quantitative.

Dichotomous characters are couplets which generally pose a true—false question. "Is the wood ring porous or is it not?" To this question there are five responses which are coded as follows:

0 = Unknown.
1 = True or first statement is correct.
2 = Variable, first statement is correct sometimes and second statement is true sometimes.
3 = False or second statement is correct.
4 = Neither statement applies.