TAXIR—A Biologically Oriented Information Retrieval System As An Aid to Plant Introduction

L. W. HUDSON, R. D. DUTTON, MARY MASSARA REYNOLDS, AND W. E. WALDEN

TAXIR, an information retrieval system designed for taxonomic data, can be effectively applied to the management of information accumulated by Plant Introduction about accessions of a major crop, *Phaseolus vulgaris*.

The Western Regional Plant Introduction Station at Pullman, Washington, is one of four similar facilities in the United States. These four stations collect and maintain records on most of the seed crops that come into the United States from various places all over the world. When a seed lot comes into the United States, it is assigned an accession number that is unique in that it applies to only one particular lot of seed. It then goes to one of the Plant Introduction Stations for increase, maintenance, and evaluation. Responsibility for the United States Department of Agriculture world collection of *Phaseolus* is assigned to the station at Pullman. Since establishment of the station in 1952, some 6,000 lines of various species of *Phaseolus* have been acquired, the great bulk of these being *vulgaris*. In addition to perpetuating these stocks, the Western Regional Plant Introduction Station is responsible for evaluating them and disseminating the assembled information to plant breeders.

The already extensive accumulation of bean germ plasm and the continually increasing evaluation notes on each accession have made meaningful access to these data difficult. At present, the description of about 35 different characteristics are recorded for each accession. Requests for information related to this collection may come from researchers anywhere in the world, and may involve simple to rather complicated questions about many characteristics of the stocks maintained. Regardless of the type of inquiry, a great deal of laborious manual searching previously was involved. A typical request may call for an early upstanding white seeded bush bean, holding its pods well off the ground, but having long pods that are circular in cross section, low in fiber, and stringless, the pods having a thick, meaty wall. Sorting for the particular combination requested through records on some 6,000 accessions, each involving 35 characteristics, is a rather awesome task. The planned inclusion of climatological and ecological data, documentary information about the origin and evaluation of disease and insect resistance, nutritional quality, genetics, etc., further compounds the problem.

Because of the difficulties involved in retrieving data, a search was initiated to locate a computer retrieval system which would be appropriate for this application. The taximetrics information retrieval system, TAXIR, was selected for testing because of its rather simple concept, its desirable space-saving features, and because the system was designed for data describing biological material. TAXIR was developed at the University of Colorado Taximetrics Laboratory by David Rogers, Henry Fleming, Robert Brill, and George Estabrook. The original version was implemented on a Control Data Corporation 6400 computer, and an improved version has just been completed (personal communication). At Washington State University a part of TAXIR was converted for operation on the IBM 360, Model 67, and this version was used for the studies de-
ACCESSION MODULE: PI(1, NAME, 2), FAMILY(2, NAME, 2),
GENUS(3, NAME, 2), SPECIES(4, NAME, 4), VARIETY(5, NAME, 7),
HYBRID(6, NAME, 3), PLOIDY(7, NAME, 5), COUNTRY(8, NAME, 6),
LOCALITY(9, NAME, 5), NAME(10, NAME, 11), PLANT HABIT(11, NAME, 3),
PLANT SIZE(12, NAME, 3),
LEAF SIZE(13, CODE, 1, 2, 3, 4, 5, 6),
PETIOL SIZE(14, CODE, 1, 2, 3, 4, 5, 6),
INTERNODE LENGTH(15, CODE, 1, 2, 3, 4, 5, 6),
PLANT ERECT(16, CODE, 1, 2, 3, 4, 5, 6),
PLANT BRANCH(17, CODE, 1, 2, 3, 4, 5, 6), PLANT STEM COLOR(18, NAME, 3),
FLOWER COLOR(19, NAME, 4),
FLOWER CONCENTRATION(20, CODE, 1, 2, 3, 4, 5, 6),
FLOWER RACEME LENGTH(21, CODE, 1, 2, 3, 4, 5, 6),
POD CONCENTRATION(22, CODE, 1, 2, 3, 4, 5, 6),
POD LENGTH(23, CODE, 1, 2, 3, 4, 5, 6, 7, 8),
POD CONSTRUCTIONS(24, CODE, 1, 2, 3, 4, 5, 6),
POD CURVE(25, CODE, 1, 2, 3, 4, 5, 6),
POD CROSS SECTION(26, CODE, 1, 2, 3, 4, 5, 6),
POD FIBER(27, CODE, 1, 2, 3, 4, 5, 6),
POD WALL THICKNESS(28, CODE, 1, 2, 3, 4, 5, 6),
POD BEAK(29, CODE, 1, 2, 3, 4, 5, 6),
POD STRING(30, NAME, 2),
POD COLOR(31, NAME, 4), POD TYPE(32, NAME, 4),
SEED COLOR(33, NAME, 6), SEED PATTERN(34, NAME, 4),
SEED SIZE(35, NAME, 4), DAYS OF MATURITY(36, NAME, 3)*

NO. OF DESCRIPTORS 36*

Fig. 1. Accession module.

scribed herein. The portion of the TAXIR system used at Washington State University is contained on about 2,200 punched cards. Approximately 6,500 additional punched cards were required to enter the data on the Phaseolus collection. When used, the information on these cards may be transferred to magnetic tape, to a data cell, or read directly by the card reader into the data storage bank of the computer. The data in the bank consists of three main groups. The first relates to the identification, i.e., accession numbers. The second describes characteristics, i.e., plant habit. The third includes the score, classification, condition, or state of the particular characteristic—i.e., for plant habit the states are bush, semi-vine, or vine. If information is missing or unknown, a special code indicating a blank is entered in the data bank. This code may be exchanged for the proper information once it has been obtained.

The TAXIR programs in use at Washington State University consist of an accession module, a define item routine, and a query routine. The accession module reads and lists the different characteristics describing each item and all the possible classes applicable to each (if numerical codes), or if a characteristic is described by name, the word NAME appears in the module. The number, letter, or symbolic codes and names make up the accession module (Fig. 1).

The accession module proved particularly useful for the detection of errors in the data lists. Many hours of tedious proofreading