



Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines

We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content

The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute

The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

BIOSYSTEMATISTS

Tuesday, October 26, 1971

SPEAKER: David J. Rogers, Professor of Biology, University of Colorado

SUBJECT: Taximetrics: Some concepts and methods.

PLACE: Social - 5:45 p. m., Wilkinson's Restaurant (John's Room)
2113 Shattuck Avenue, Berkeley. Contribution: 75 cents.

Dinner - 6:30 p. m., Wilkinson's Restaurant. Minimum meal
about \$2.50 plus tax and tip.

Lecture - 7:45 p. m., 2507 Life Sciences Building, University
of California, Berkeley. (Enter through door near
Biology Library).

Room 2000 Life Sciences Bldg

Came 3:30
from FRI

w/ Dr. Wiggins

Introductory comments:

1. What I hope to do here is present in very general terms what taximetrics is all about. Since I expect that each of you comes with differing interests and backgrounds, it is impossible to speak more specifically about any of the topics. For anyone interested in further details, I have provided an annotated list of our ~~xxxx~~ publications which may help further your understanding.

humor.

2. Derivation of the word, taximetrics.--and Ernst Mayr's heavy German /

3. Difference from numerical taxonomy--the "fly paper gambit"--get

close to it and you're stuck.

Sokal, and his "operationalism"--change the science to his image of i

The "intelligent ignoramus" a la Sokal is an insult.

By contrast, taximetrics aims to emulate existing good taxonomic practise, past and present. Taximetric methods should aid the taxonomist not substitute for him.

Edgar Anderson provided basic impetus to my work.
method of description and the hybrid index.

5. Must say something about the aims and objectives of taxonomy.

- a. from a biological point of view
 1. to discover relationships of organisms past and present-- the phylogenetic approach.
 2. to discover the biological species as the basic unit of life
- b. from an informational point of view
 1. a precise and efficient means to organize information about the organic world.
 2. an information retrieval system -- the efficiency of an herbarium or zoological museum.

c) From a use-oriented point of view--to make possible the means by which plant and animal breeders know which plants to look for, where they may find them, in what type of ~~taxonomy~~ ecological niche.

What separates taximetrics from, say, biosystematic or chemotaxonomic studies?

1. Taximetrics examines the ways of thinking about our work, or the way we "do" taxonomy. In a sense, it is methodology, emulating the procedures used in deciding that two specimens (or preferably, objects) are similar or different, that two species are "closely related", or not, how ~~tax~~characters are good or not, etc.

2. In my approach, taximetrics procedures are developed using a team approach. Appropriately trained mathematicians, good programmers (professional) work together with a trained taxonomist (or preferably, several taxonomists). After overcoming initial communications ^{difficulties} ~~problems~~ (establishing common vocabularies), problems can be more meaningfully attacked.

out; } 3. Following development of a set of procedures, these are tested on a variety of taxonomic problems before being released to public examination.

The importance of mathematics.

What sorts of math? Taxonomists work in a finite world--their math, therefore, must be of a finite sort, not infinite. Taxonomists are interested in the syntheses of data to derive meaning, and the math which supports these syntheses must be used, rather than analytical (at least in the final stages).

Taxonomists use a mix of both qualitative and quantitative information to derive their classifications, depending preponderantly on qualitative data (even though quantification can be made, it does not necessarily add accuracy. Consider colors of flowers--how to you quantify these? Therefore, the math must be capable of using both qual & quan together. The math, therefore, falls under the general term "combinatorics", and is relatively modern. I suggest you read Estabrook's paper for further details. In general such math as graph theory, information theory, lattice theory, partitions and bloc designs, partial orders, and certain algebras are particularly powerful allies for taxonomic work.

Math--continued

The above implies that we let our problems in taxonomy be the guide to the math which is appropriate, rather than insisting that the problems is one of statistics, or algebra, or calculus, or whatever. The taxonomist must use his ability to key things out to get the right sort of mathematician and mathematics.

Computation--the application of various sorts of computers.

Taxonomists employ heavy loads of data, though they are of relatively and simple similar/types, and these require very large memories in computing machines. The efficient manipulation of the memory systems of the computer must be one of the most important considerations for the programmer of taxonomic procedures.

The math computations made for our programs are relatively simple, and do not require most of the computational ~~skill~~ capacity of the computer (although this is what a computing machine^{does} is best).

A machine designed to our specifications would look quite different from one for, say, a theoretical physicist.

How do you decide where to start in?

A break-down of taxonomy--the flow charts.

early computer programs

How nearly all ~~of us~~ started--that part of classification frequently referred to as "speciation"--clustering.

Once one has a good clustering model, one then discovers that the

the characters and their attributes
input information/to the clustering model may be faulty, and the

next step is the design of a procedure to test the values of ~~ixix~~

the characters we have chosen.

If one is of a theoretical frame of mind, the next logical step

is to know how to choose the characters in the first place--we

have not attacked this last one at all.

Or one may choose, as we did, to work on a system to keep track

of the thousands of details needed to do taxonomic work--a record

keeping system with instant recall.

The existing taximetrics programs.

Three major programs, each with a large number of sub-programs, are

now operative. Each has stood the test of actual taxonomic testing

with a large number of different problems. These programs are:

TAXIR-- Taxonomic Information Retrieval. This powerful system is our latest development, funded by a \$300,000 grant from NSF. It provides a means by which the taxonomist can use his own natural language (or a close proximity thereto) to store information in any form most convenient to him, to retrieve the information in any combination of importance to him. It has all the required capacities ~~of~~ necessary in any system, such as correction of erroneous data, deletion of unnecessary (or redundant) data, addition and updating, ordering of tabular information, giving precise answers to a question or providing a complete listing of all the data in the system.

We have used the system to prepare indexes to exsiccate, species ~~d~~istribution data, bibliographic information, type specimen lists, ecological and other habitat data for each species, and to aid in the making of keys to the species. Various other applications have been made--this system was chosen by the consortium of Universities along the Gulf of Mexico as their information retrieval system for the Gulf Oceanographic Decade study, just now under way.

CHANAL--Character Analysis. This program is extremely powerful ally

to the taxonomist in deciding which characters are the most powerful

for the purposes of making a classification. * The mathematical

base for this program is information theory. The character is

considered to be "information" about the set of objects under study.

(what does include?)
As an example of the difference between a methodological and a biological approach, a character is defined as a single basis for comparison defined over the objects under study. This definition assumes that the taxonomist is sufficiently knowledgeable of biology not to be misled by problems of homology, and that he is capable of applying comparative methods. It has the methodological benefits of being sufficiently general that it does not get mixed up with biological problems involved in the genetic constitution of different characters.

each character shares in common with each other character, the

redundancy of this information, and for each object, which character

do you explain definition

state of a character is most common, next most, and least common.

This program helps the taxonomist decide whether the information

he has chosen will be most useful in making the classification.

It has many other hints, too detailed to be included in this

discussion.

variable

GRAPH--Graph similarity and clustering program.

This model is based upon three widely accepted taxonomic principles:

1. A classification for a collection of objects is a hierarchical, two-dimensional partitioning of the objects. A partition of the objects is a subdivision of the collection into sub-collections, such that each object is in one and only one sub-collection.
2. For any given partition, two objects at least as similar as the degree considered for this partition should not be placed into different sub-collections.
3. The sub-collections of a given partition should be isolated from one another; that is, there should be some discontinuity between the members of different sub-collections, found in the information about the objects.

Theoretically, none of these principles has to hold for a classification, but in practise, taxonomists have used these as their guiding lights time after time. This is one reason we say that we have followed taxonomic practise in development of our models.

From these three principles, we derived a model based on the mathematics called graph theory. I could give greater details on this branch, but let it rest by saying that it is a very useful form

for taxonomic work.

The computer program for GRAPH contains two major components, the first being a derivation of similarity measures for each object in a study with every other object in the study. The second part is the clustering program, using as input the similarity measures derived in the first part. The objective of the program is to provide a complete story on how each object is related to each other object (according to the defining characters), and how the pairs of objects are clustered into larger groupings. At the start of the program, each object is a cluster by itself, and by progressive stages, clusters are formed which at the end of the program have all been placed into one large cluster.

The program is designed as an aid to the taxonomist in making decisions about the taxa in his classification. It aids by suggesting relationships that may not have occurred to him, by giving progressive information about each specimen, and several measures of relatedness.

SOME READINGS IN TAXIMETRICS

GENERAL

1. Rogers, D. J. 1963. Taximetrics--new name, old concept. *Brittonia* 15: 285-290.
Where "taximetrics" was first used and defined.
2. Rogers, D. J., H. S. Fleming and G. Estabrook. 1967. Use of computers in studies of taxonomy and evolution. In *Evolutionary Biology*, Vol. I, 169-196. Eds. Dobzhansky, T., M. K. Hecht, and W. C. Steere. Appleton-Century-Crofts.
This paper gives the general background for development of taximetrics, although the title is misleading. We give some historical background, rationale in mathematics, descriptions of computers and programming, and the various categories in evolutionary and taxonomic work where computers had been applied up to that time.
3. Rogers, D. J. and S. G. Appan. 1969. Taximetric methods for delimiting biological species. *Taxon* 18: 609-624.
Herein, we describe the integrated use of all of our methods to aid in the study of biological species. The methods themselves do not delimit the biological species, but point to the step-wise procedures which, if followed, are powerful techniques in guiding the discovery (or delimitation).

TAXIR

1. Estabrook, G. F. and R. C. Brill. 1969. The theory of the TAXIR accessioner. *Math. Biosciences*. 5: 327-340.
This paper gives the mathematical basis for the TAXIR Information retrieval system. A powerful method to store the information, address the information, and retrieve it is described.
2. Rogers, D. J. 1970. Theoretical and practical considerations on data structuring for a computerized information retrieval system. In *Archeologie et Calculateurs* (Marseille). pp. 145-159. Editions du Centre National de la Recherche Scientifique. 15, Quai Anatole-France. Paris VII^e
I compare the information systems employed by systematic curators with the model we developed, and indicate some of the methods to prepare data for storage.

CHANAL

1. Hawksworth, F. G., G. F. Estabrook and D. J. Rogers. 1968. Application of an information theory model for character analysis in the genus Arceuthobium. (Viscaceae). *Taxon* 17: 605-619.
The methods of use of the character analysis program are described in biological terms, and the application of the findings made for the value of the characters to be employed in the classification of Arceuthobium.

TAXONOMIC METHODS IN TAXIMETRICS

2. Estabrook, G. F. 1967. An information theory model for character analysis. *Taxon* 16: 86-97. The mathematical arguments for this program are given.
- GENERAL
3. Fleming, H. S. and D. J. Rogers. 1970. A classification of Manihot esculenta using the information carrying content of a character as a measure of its classificatory rank. *Proc. Int. Symp. Trop. Root Crops (Hawaii)*. pp. 66-70. This interesting application of the character analysis program indicates that it provides another method of classification, in addition to the power as a character analytic technique.
- USE OF COMPUTERS
1. Estabrook, G. F. 1967. Use of computers in studies of taxonomy and evolution. In *Evolutionary Biology*, Vol. 1, 169-196. Eds. Dobzhansky, T., V. T. Becht, and W. C. Steara. Appleton-Century-Crofts.
- GRAPH
1. Wirth, M., G. F. Estabrook and D. J. Rogers. 1966. A graph theory model for systematic biology, with an example for the Oncidiinae (Orchidaceae), and Systematic Zoology 15: 59-69. This paper describes in taxonomic terms the clustering methods developed in the Taximetrics Lab. The paper does not attempt to deal with the mathematical arguments.
 2. Estabrook, G. F. and D. J. Rogers. 1966. A general method of taxonomic description for a computed similarity measure. *Bioscience* 16: 789-793. This paper describes the important process for structuring characters to be used in classification.
 3. Estabrook, G. 1966. A mathematical model in graph theory for biological classification. *Jour. Theor. Bio.* 12: 297-310.
 4. Irwin, H. S. and D. J. Rogers. 1967. Monographic studies in Cassia (Leguminosae-Caesalpinioideae). II. A taximetric study of section Apoucouita. *Mem. N.Y. Bot. Gard.* 16: 71-118. This is one of our first applications of the GRAPH clustering model. The section Apoucouita of Cassia contained some of the most complex classificatory problems in this large genus, and the computer model proved of great assistance in sorting out the specific and sub-specific taxa.
 5. Prance, G. T., D. J. Rogers and F. White. 1968. A taximetric study of an angiosperm family: generic delimitation in the Chrysobalanaceae. *New Phytol.* 68: 1203-1234. Four computer methods were compared for their capacity to make classifications, and our GRAPH model came out on top. In this case, we were comparing genera within a family.
 6. Arp, G. and D. J. Rogers. 1970. A computer-aided classification of the varieties of Pediocactus simpsonii (Engelm.) B. & R. *Cactus & Succ. Jour.* 42: 40-43.
 7. Stearn, W. T. 1971. A survey of the tropical general Oplonia and Psilanthele (Acanthaceae). *Bull. Br. Mus. (Nat. Hist.) Botany*. IV (7): 259-323, plates 38-47. Stearn used our GRAPH model to aid in this study, particularly with the genus Oplonia.

8. Rogers, D. J. and H. S. Fleming. 1972 (In press) Economic Botany. A monograph of Manihot esculenta. The GRAPH model was our chief method to sort out the complex relationships within a cultivated species. The methods of use, and several suggestions on the intraspecific problems of cultivated plant taxonomy are described in sufficient detail so that workers can fully grasp our concepts. Proc. Int. Symp. Trop. Root Crops (Hawaii), pp. 66-70. This interesting application of the character analysis program indicates that it provides another method of classification, in addition to the power as a character analytic technique.

GRAPH

1. Wirth, M., C. F. Estabrook and D. J. Rogers. 1965. A graph theory model for systematic biology, with an example for the Oncidiinae (Orchidaceae). Systematic Zoology 15: 59-69. This paper describes in taxonomic terms the clustering methods developed in the Taximetrix Lab. The paper does not attempt to deal with the mathematical arguments.
2. Estabrook, C. F. and D. J. Rogers. 1966. A general method of taxonomic description for a computed similarity measure. Bioscience 16: 789-793. This paper describes the important process for structuring characters to be used in classification.
3. Estabrook, C. 1966. A mathematical model in graph theory for biological classification. Jour. Theor. Bio. 12: 297-310.
4. Irwin, H. S. and D. J. Rogers. 1967. Monographic studies in Cassia (Leguminosae-Caesalpinioideae). II. A taximetric study of section Apoucouita. Mem. N.Y. Bot. Gard. 16: 71-118. This is one of our first applications of the GRAPH clustering model. The section Apoucouita of Cassia contained some of the most complex classificatory problems in this large genus, and the computer model proved of great assistance in sorting out the specific and sub-specific taxa.
5. France, C. T., D. J. Rogers and F. White. A taximetric study of an angiosperm family: generic delimitation in the Chrysobalanaceae. New Phytol. 68: 1203-1234. Four computer methods were compared for their capacity to make classifications, and our GRAPH model came out on top. In this case, we were comparing genera within a family.
6. Arp, G. and D. J. Rogers. 1970. A computer-aided classification of the varieties of Pediocactus simpsonii (Engelm.) B. & R. Cactus & Succ. Jour. 42: 40-43. An intraspecific study of a cactus species.
7. Stearn, W. T. 1971. A survey of the tropical general Oplonia and Palaeanthea (Acanthaceae). Bull. Br. Mus. (Nat. Hist.) Botany. IV (7): 259-323, plates 38-47. Stearn used our GRAPH model to aid in this study, particularly with the genus Oplonia.