



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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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.

TOPICAL OUTLINE

1. General nature of classification

Theme: Classification requires non-overlapping subsets. Descriptors serve function of assignment. Choice of descriptors is "arbitrary"; that is, dependent on usefulness to the classifier. Effect of independence of descriptors and of large numbers of descriptors. Classification in the general sense is related to problem of information storage and retrieval.

Method: Possibly start with a collection of "nonsense" objects to classify of the idea of library classification.

Ref. : Sokal and Sneath - chapter on phyletic vs. phenetic

2. Particular problems of biological classification

Theme: It is inevitable that biological classification is pervaded by the evolutionary idea that organisms did "in fact" arise in one particular order. This could be represented as a tree diagram if we but knew the actual relationships. Question is should classification incorporate evidence (hard or conjectural) on relationships or should the relationships be deduced from the classification. At the very least both approaches should not be applied to one classification. Thus the dimension of classification for detection of evolutionary relationships is added as a possible purpose in addition to information retrieval and storage; the problems of cladistics must be considered.

Ref.: Estabrook - "Cladistic Methodology: A Discussion of the Theoretical Basis for the Induction of Evolutionary History." Annual Review of Ecology and Systematics, V.3, 1972.

not used

3. Hierarchical ordering in classification

Theme: There are two fundamental problems here. One is to define the distances between and among supraspecific groups and the other is to delimit the species itself.

Ref.: Erlich, P.R. 1958. Problems of higher classification. Syst. Zool. 7:180 - 184.

Fitch and Margoliash, 1967. Construction of Phylogenetic Trees. Science 155:279-84.

Estabrook, above on distance measures.
Rogers and Appan - Taximetric Methods for Delimiting Biological Species (reprint from Taxon, 1969)

Hutchinson, G.E. When are Species Necessary?, in Population Biology and Evolution, Lewontin, ed. 1963

TOPICAL OUTLINE

4. The Taximetrics method as an aid to defining and implementing our thinking processes in the classification procedure in taxonomy and other disciplines.

Theme: "Aim is to amplify and objectify traditional methods . . . not . . . substitute for the specialist but aid him in exercising his intuitive faculties astutely."

All scientific investigation is based on observational data - often in large quantities and without obvious relationships. Scientific knowledge from this data must result from analysis.

Techniques:

1. Careful consideration of purpose in operational terms.
2. Flow chart plan of exact procedure to achieve the goal.
3. Data storage in a fashion which allows for necessary manipulation to suggest interrelationships as well as for retrieval.
4. Collation and comparison of many samples over many characters with some means of determining relative "worth" of the various observations.
5. Determination of similarities among groups and distances between groups.

Method:

After beginning with the overall description in delimiting biological species paper above, present several examples of the process as applies to taxonomic studies by asking students or groups of students to read and report on one of the following examples.

Then assign as a term project flow charting procedures and purposes in their own discipline.

Ref.:

Rogers and Tanimoto, A Computer Program for Classifying Plants. Science 132:1115-1118, 1960

Hawksworth, Estabrook, and Rogers, Application of an information theory model for character analysis in the genus *Arceuthobium* (Muscaceae) Taxon 17:605-619, 1968.

Legendre, Schreck, and Behnke, Taximetric Analysis of Selected groups of western North American *Salmo* with respect to phylogenetic divergences. Syst. Zool. 21:292-307, 1972.

With, Estabrook, and Rogers, A Graph model for systematic biology with an example from the *Oncidiinae* (Orchidaceae). Syst. Zool. 15: 59-69, 1966

Rogers - (classification of *Manihot*) Economic Botany 27# 1, 1973
also Flora Neotropica Monograph No. 13 (1973)

Hull, L.D. The Operational Imperative - sense and nonsense in operationalism. Syst. Zool. 16:438-57, 1968

TOPICAL OUTLINE

5. Taximetrics as compared with "classic" numeric taxonomy

Theme: The initial difference in approach involved the issue of whether numeric, computer aided methods should attempt to aid or supercede the judgement of the classic taxonomist. Numeric taxonomists rely much more on parametric statistics and its accompanying assumptions on normal distributions.

Ref.: Legendre talk at NT meeting in August 1973.

Sneath and Sokal - new edition.

Sokal & Sneath (1970) *The Intelligent Ignoramus, an experiment in numerical taxonomy*, *Taxon* 19(3): 305-319.

6. Introduction to the computer - its strengths and limitations.

Theme: In general the computer can do only a few simple procedures such as mathematical processes, comparisons, and simple choices. Its usefulness lies in the fact that it does these things very fast. The computer program is a set of very exactly stated instructions written in a computer language.

Techniques: key punching, control cards, getting on line, data cards, sub-accounting details, tape requests and special outputs.

Method: Spend one period in the computer terminal getting acquainted with above techniques and with sources of aid and assistance over there.

Ref.: 7.

Rogers, Fleming, and Estabrook - Use of Computers in Studies of Taxonomy and Evolution. from *Evolutionary Biology*, V. 1, Dobzhansky editor, 1967.

7. The Mathematical Basis for Classification

Since classification is division into non-overlapping subsets, its theory is based on set theory and Boolean algebra. An introduction to the terminology and concepts of this powerful technique is useful. Due to the fact that the fundamental units of biology are non-identical, finite methods are more applicable than the analytic methods that have been so successful in the physical sciences.

Ref.:

Kemeny and Snell, *Finite Structures in Mathematics and Introduction to Finite Mathematics*

Elsasser - *Physical Foundations of Biology*

TOPICAL OUTLINE

8. Nature of Descriptors and Characters and their States, and of Objects

Theme: Descriptors and be any item of information which applies to an object - that is a sample or item upon which data are collected. The descriptor may exist in any of one to several states - as blue is a state of the descriptor color.

Characters are defined rather more thoughtfully in that which characters one chooses reflects the biologist's judgement as to the meaningful relation of the observation to his system and to his purposes. Further the character states must reflect in an honest way the distribution of that character in his sample.

A descriptor or character by item or object matrix may be built up to clarify the concepts of descriptor, state, and item.

Method: By the time we get to this point students should have their projects chosen and will be asked to turn in a descriptor by item matrix or table of their data.

9. Introduction to TAXIR

Theme: Develop the idea that any information involving comparable data for a number of objects can be stored in the system. Further that the query procedure gives enormous opportunity for trying out any number of possible relationships among the data.

Give examples of its use. Then present the technical details of how the program works and how cards must be punched etc. to use it. Details of TELEX use.

Method: The Christmas Bird count project as an example of the use of the program in a very large set of data.

Ref.:

Rogers - Theoretical and Practical Considerations on Data Structuring for a Computerized Information Retrieval System

"Manuals" for using TAXIR developed by last year's class.

Users manual for the new program as now available in computer center.

Assignment - Apply TAXIR to their data and query the data bank by card and by TELEX if appropriate.

10. The Structure of Characters - Determining their Relative "Worth"

Theme: Classically this¹⁵ the question of weighting and involves the more fundamental and philosophical question of the role of judgement on the part of the systematist who has wide experience with a group of organisms. The object of any computer assisted analysis of characters is to dissect out elements of the process of judgement used intuitively by the scientist and apply them systematically to all possible pairs of a set of proposed characters. CHARANAL does this using the information theory; there are other more classically statistical means for approaching it.

Ref. - Sneath and Sokal - chapter on weighting problem and statistical approaches to it.

Hawksworth, Estabrook, and Rogers, Application of an Information theory model for character analysis in the Genus *Arceuthobium*.

11. Information Theory and its Application to Biology

Theme: Information theory is based on the entropy concept of classical thermodynamics. The information function measures the amount of uncertainty in a system which indicates the amount of information which is potential, in a sense, if it can be extracted from the system. CHARANAL applies this measure to pairs of characters. It is interesting to note that this function turns up in other biological situations -- for example, it is used as a measure of diversity in ecology.

Ref.:

Estabrook, G. An information theory model for character analysis, *Taxon* 16:86-97. 1967

Legendre, P. and Rogers, D.J. Characters and Clustering in Taxonomy: A Synthesis of Two Taxometric Procedures. (reprint - 1st section)

Morowitz - Entropy for Biologists

12. The CHARANAL Program -- Procedures for its Use and Interpretation

Theme: Details of preparation of data, cards etc. for running the program. An example of output will be used to demonstrate the expected output and to interpret the results.

Ref.: CHARANAL manual

Assign.: run their data on CHRRANAL and interpret results

13. Similarity Measures

Theme: Classification depends on the degree of similarity of the objects first, and then on the distance between groups of object with a given degree of similarity; that is, "clusters" are arranged in a "hierarchy". (Different studies may be more interested in the former than the latter and vice-versa.). This points to the need for devising some means of quantifying similarity. Once this is done a computer approach to their arrangement so as to make maximum use of the similarities of all objects in the system is an obvious need.

Ref.

Estabrook and Rogers, A General Method of Taxonomic Description for a Computed Similarity Measure. (reprint)

Estabrook, A Mathematical model in graph theory for biological classification, *J. Theoret. Biology* 12:297-310

Wirth, Estabrook, and Rogers, A Graph theory model for systematic biology, with an example for the Oncidiinae.

Sneath and Sokal - on statistical measures.

14. Using and Interpreting the GRAPH program

Theme: same as for use of CHARANAL

Ref.: Cluster Analysis Program manual

Assign. : run data on program and interpret results.

15. Systema Analysis in Biology and the Future of Taximetrics

Theme: further developments can lead to "decision making" assistance through computer analysis of the data of complex systems.

Method: What I would like to do is to get Gil Hersch to come in and talk in this area.

The final exam would consist of reports by students on their projects - oral presentations as we did last year.