



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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

ACM REFERENCE GUIDES

SPECIAL INTEREST GROUPS / SPECIAL INTEREST COMMITTEES (AS OF JULY 15, 1967)

SIGIR INFORMATION RETRIEVAL

application of machines to the storage, retrieval and dissemination of information / information retrieval theory, programming, file preparation, searching strategy, output schemes, IR system evaluation, development of equipment best suited for these tasks / chairman: herbert ohlman, xerox corporation, box 150, rochester, new york 14604 / annual dues \$2 / SIGIR Forum (informal newsletter)

SIGCPR COMPUTER PERSONNEL RESEARCH

investigation of the nature of computer personnel and their jobs, identification of common needs among computer personnel, research on related problems / chairman: robert a. dickmann, johns hopkins university, applied physics laboratory, 8621 georgia avenue, silver spring, maryland 20910 / annual dues: ACM members \$2, non ACM members \$4 / annual meeting proceedings

SIGUCC UNIVERSITY COMPUTING CENTERS

discussions of mutual problems / guidance to colleges and universities desiring to establish or reorganize computing centers / chairman: m. mulder, computing center, university of wisconsin, madison, wisconsin 53705 / annual dues \$1 / SIGUCC Newsletter

• SIG IC FUNCTIONS

Info: on processing comprises many fields, and continually evolves new subsectors. Within ACM these receive appropriate attention through Special Interest Groups (SIGs) and Special Interest Committees (SICs) that function as centralizing bodies for those of like technical interests . . . arranging meetings, issuing bulletins, and acting as both repositories and clearing houses. ACM members may, of course, join more than one special interest body. The existence of SIGs and SICs offers the individual member all the advantages of a homogeneous narrower purpose group within a large crossfield society.

National meetings provide an opportunity for SIGs and SICs to sponsor technical sessions in their areas of interest, as well as an opportunity for administrative meetings of each group and committee.

There is no distinction in purposes or goals between a SIC and a SIG. The only difference is one of organization. Two organizational forms are provided to accommodate differences in the number and administrative interest of the members. Also, the committee form may be used as an interim form before the organization of a group.

• SIG/SIC MEMBERSHIP

To join a SIG or SIC, ACM members should contact the particular chairman listed.

For information regarding the formation of SIGs and SICs, contact the Chairman of the ACM Committee on Special Interest Groups and Committees: C. L. Bradshaw, MSFC—NASA, Bldg. 4663, Huntsville, Alabama.

SIGMAP MATHEMATICAL PROGRAMMING

application of computing machinery to the many phases of the problem of constraint optimization / includes linear programming and its extensions to convex programming, integer programming and network flow / included but with less attention are dynamic programming and some disciplines ordinarily grouped under the heading "constrained optimization" / chairman: m. a. eryomson, esso research and engineering company, p.o. box 101, florham park, new jersey 07932 / annual dues \$2 / SIGMAP Newsletter

SICART ARTIFICIAL INTELLIGENCE

creation of programs, machines and other human artifacts which learn from experience or exist in other forms of intelligent behavior / chairman: rebecca prather, system development corporation, santa monica, california 90406 / SICART Newsletter

SICBDP BUSINESS DATA PROCESSING

exchange on problems of mutual interest / current emphasis: establishing local SIGBDP's in conjunction with local acm chapter / chairman: george glaser, mckinsey and company, inc., 100 california street, san francisco, california

(SIC) BIO BIOMEDICAL INFORMATION PROCESSING

digital computer codes, programs, routines, techniques applicable to biomedical problems / facilitates biomedical computer programming tasks impractical to unorganized units / defines standards of conventions / considers standards for equipment relating digital computers to biomedical laboratory equipment / stimulates better understanding of potentialities of digital computers in biosciences / chairman: r. w. stacy, university of north carolina, department of surgery, chapel hill, north carolina 27514

SICCAP CIVIL ENGINEERING, ARCHITECTURE AND PLANNING

disciplines concerned with physical solutions of problems of environmental control, especially civil engineering, architecture and (city, regional and transportation) planning; applications to management functions, technical analysis, system design / liaison with members of applications-oriented professional societies in these areas / chairman: allen herschdorfer, department of civil engineering, massachusetts institute of technology, cambridge, massachusetts 02139

SICCSIM COMPUTER SYSTEMS INSTALLATION MANAGEMENT

new installation management concepts; selection, installation and operation of electronic components, computing centers / chairman: j. r. buckingham, c/o acm headquarters, 211 east 43 street, new york, new york 10017

SICDA DESIGN AUTOMATION

computer programs and techniques in design, manufacturing and testing of electronic components, including packaging standards, printing and checking of logic diagrams, optimum methods of back panel wiring and diagnostic and test specifications; use of cathode ray tubes, light pens, automatic drafting equipment and other display equipment in conjunction with computers / chairman: m. weindling, douglas aircraft, space systems center, huntington beach, california

SIGGRAPH GRAPHICAL DISPLAY

all aspects of graphical man/machine communication: hardware, languages and data structure, methodology and applications / chairman: sam matas, 590 madison avenue, new york, new york 10022

SICNUM NUMERICAL MATHEMATICS

serves the area interfacing between mathematics and computing / currently, an investigation of testing and certification techniques for numerical subroutines / chairman: j. f. traub, computer science research center, bell telephone laboratories, murray hill, new jersey 07971 / SICNUM Newsletter

SICODCPT DIGITAL COMPUTER PROGRAMMER TRAINING

Investigation of current problems, development of improved methods, exploration of new areas / chairman: gloria m. silvern, education and training consultants, 979 teakwood road, los angeles, california 90049 / announcements

SICPLAN PROGRAMMING LANGUAGES

theory, design, implementation, description and application of computer programming languages that permit the specification of a variety of different computations, thereby providing the user with significant control over the computer's operation / activities directed to the specialist / chairman: christopher j. shaw, system development corporation, santa monica 90406 / SICPLAN Notices, monthly

SICRTP REAL TIME PROCESSING

programming techniques associated with real time processing, as distinguished from non-real-time processing / current emphasis: bibliography, glossary, history, and state of the art / chairman: sherman blumenthal, union carbide corporation, 270 park ave., new york, n.y. 10017

SICSAM SYMBOLIC AND ALGEBRAIC MANIPULATION

design, usage, and applications of systems for symbolic and algebraic manipulations—list processing systems, string handling systems, formula manipulation systems / language comparison and applications subcommittees are currently active / chairman: jean a. sammet, IBM corporation, 545 technology square, cambridge, massachusetts 02139 / SICSAM Bulletin (informal newsletter)

(SIC)² SOCIAL IMPLICATIONS OF COMPUTING

a forum for discussions related to the social responsibilities of computer professionals / bibliography service in the form of edge-notched cards—in the launching stage / chairman: anatole holt, applied data research, route 206 center, princeton, new jersey 08540

SICTIME TIME SHARING

computer utilization in multiprocessing, multiprocessing environments, particularly for remote, multiuser interactive applications / chairman: a. m. rosenberg, scientific data systems, 1649 17th street, santa monica, california 90404 / SICTIME Bulletin (informal newsletter)

SICUDS URBAN DATA SYSTEMS

state and local government systems—data base management, regional planning, law enforcement / chairman: john a. postley, informatics inc., 5430 van nuys blvd., sherman oaks, california 91403

3 November 1967

Mr. Herbert Ohlman
Xerox Corporation, Box 150
Rochester, New York 14604

Dear Mr. Ohlman:

Please add my name to the members of the SIGR. I am
enclosing \$2 for the annual dues.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm
Enc.: \$2

2 February 1968

V

NOV 14 1967

UNIVERSITY OF MARYLAND
COLLEGE PARK 20740

COLLEGE OF AGRICULTURE
DEPARTMENT OF BOTANY

November 8, 1967

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dave:

This is just a note to thank you for your kindness in sending the batch of reprints on numerical taxonomy and related subjects. Tomorrow our seminar participant is going to hold forth on the subject and I am looking forward to her presentation with great interest. If you ever have an opportunity to visit this area, we could round up enough interested people for you to talk at. Will you let me know if you'd like a captive audience sometime? At least our graduate students will have had a little exposure by that time. Thanks again for your generosity.

Sincerely yours,



William L. Stern
Professor of Botany

dc

I suppose that this project should be talked over with Rodock and many others at the museum here before it gets to the point where we should do something positive about a proposal to NSF. I will be glad to participate in the discussion about the sort of work provided there is somebody in the museum willing to go along with it. I do not want to work with this thing if the people in the museum are less than enthusiastic. I know Bill Weber is enthusiastic but can say nothing about the others (Rodock, et al.).

11 October 1967

Dean E. James Archer
Regent 201
University of Colorado
Boulder

Dear Jim:

The opportunity to do some collaborative research between the museum here and the Smithsonian Institution is now a very live option. President Johnson recently pointed out that museums were in need of assistance to do their job, and it has been an "in" thing to do to work on information retrieval. Bill Weber's idea and mine was to make some sort of communication system via teletypewriter or some other line such that data could be exchanged between the Smithsonian and our museum. The Smithsonian has established a number of procedures for data recording with automatic input devices (similar to the Flexi-writer) which save immense amounts of time in terms of secretarial and curatorial activities. Our museum would benefit by the knowledge in the Smithsonian already available for processing museum-type information and this tie between our museum and the Smithsonian could be thought of as a prototype system for many other museums around the country. Information retrieval could work very nicely in a situation where queries could be made between large and small institutions concerning the disposition of certain specimens or the location of rare specimens or the fossil record of a certain area, etc. The objective of such a tie could be to make more efficient the type of work for which museums are particularly well suited.

I think the best people to talk with about this particular type of activity are Dr. Burton Adkinson, Head of the Office of Scientific Information Services, National Science Foundation, and Dr. Sydney Geller, Assistant Secretary for Science of the Smithsonian. Adkinson, I am sure, would be well disposed to the development of a system. He has encouraged biologists to apply to NSF for information retrieval work. I know that Geller and others at the Smithsonian will be more than glad to collaborate on this experimental phase, because I have talked with him and others (Dr. Don Squires, Deputy Director, Natural History Museum) there.

I suppose that this project should be talked over with Rodeck and many others at the museum here before it gets to the point where we should do something positive about a proposal to NSF. I will be glad to participate in the discussions about this sort of work provided there is somebody in the museum willing to go along with it. I do not want to work with this thing if the people in the museum are less than enthusiastic. I know Bill Weber is enthusiastic but can say nothing about the others (Rodeck, et al.).

Dean E. James Archer
Room 201
University of Colorado
Boulder

Dear Jim:

The opportunity to do some collaborative research between the museum here and the Smithsonian Institution is now a very live option. President Johnson recently pointed out that museums were in need of assistance to do their job, and it has been an "in" thing to do to work on information retrieval. Bill Weber's idea and mine was to make some sort of communication system via teletypewriter or some other line such that data could be exchanged between the Smithsonian and our museum. The Smithsonian has established a number of procedures for data recording with automatic input devices (similar to the Elex-Writer) which save immense amounts of time in terms of secretarial and curatorial activities. Our museum would benefit by the knowledge in the Smithsonian already available for processing museum-type information and this tie between our museum and the Smithsonian could be thought of as a prototype system for many other museums around the country. Information retrieval could work very nicely in a situation where queries could be made between large and small institutions concerning the disposition of certain specimens or the location of rare specimens or the fossil record of a certain area, etc. The objective of such a tie could be to make more efficient the type of work for which museums are particularly well suited.

I think the best people to talk with about this particular type of activity are Dr. Burton Aukman, Head of the Office of Scientific Information Services, National Science Foundation, and Dr. Sydney Galter, Assistant Secretary for Science of the Smithsonian. Aukman is very well disposed to the development of a system. He has encouraged biologists to apply to NSF for information retrieval work. I know that Galter and others at the Smithsonian will be more than glad to collaborate on this experimental phase. Because I have talked with him and others (Dr. Don Spurr, Deputy Director, Natural History Museum) there,

16 October 1967

Dr. William L. Stern
Department of Botany
College of Agriculture
University of Maryland
College Park, Maryland 20740

Dear Bill:

A complete set of reprints was sent out recently. Some of the papers are of historical interest only - namely the Science paper of 1960 and the earlier BioScience paper by Fleming and me.

I cheer your idea of methodology in plant taxonomy and I surely hope that you will bear down hard on the business of structuring of characters. Whether one uses the computer or not, the whole taxonomic process rises or falls on the basis of the good construction of the input data. I wish that we could give you some sort of seminar ourselves to go along with the paper from BioScience by Estabrook and me because, while that says all we want to say, it is rather straight off the shoulder, and as a result could stand some additional explanation. However, take a good look and let me know if there is anything else we can do to help your seminar. I will be happy to do so.

Incidentally, there is on the staff of the University of Maryland another escapee from the Washington bureaucracy, namely Dr. Jerry Kidd who is interested in information retrieval. I recommend that you have a chat with him sometime, because systematic botanists have great need for retrieval of data and Jerry is very good about this sort of thing.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

UNIVERSITY OF MARYLAND
COLLEGE PARK 20740COLLEGE OF AGRICULTURE
DEPARTMENT OF BOTANY

October 3, 1967

Dr. David J. Rogers
Department of Biology
University of Colorado
Boulder, Colorado

Dear Dave:

This semester I am holding a graduate seminar entitled, "Modern Methodology in Plant Taxonomy." We will be discussing the various "new" chemical techniques as well as some of those in numerical and computer taxonomy. To these ends, I have assigned participants to various topics and handed out a list of references to give the students a start on their bibliographic searches.

One of the girls turned up your recent article with Estabrook in BioScience (November 1966) at the end of which are references to several other papers which you have done alone or in conjunction with others. Our library at the University of Maryland is not rich in taxonomic periodicals (a situation which I am striving to remedy as fast as possible) and it would be of considerable advantage to us if you could send me a selection of reprints of your papers on numerical and computer taxonomy. Particularly, we would like copies of the BioScience paper and your papers in Science (132: 1115-1118) and in Systematic Zoology (15: 59-69).

I have told my students not to write to authors for reprints but to let me do this for them. In this way I can keep the reprints and make them available to students each semester. This will keep them from pestering authors time and again.

Very best personal regards.

Sincerely yours,

William L. Stern
Professor of Botany

dc

Mailed Oct. 12, 1967

Use of Computers - Evol. Biology
A Math. Model In Graph Theory
An Information Theory Model
Graph Theory Model for Syst. Biol.
Computer Program for Classifying Plants
A General Method of Taxonomic Descript. BioScience

Put him on our

PLYMOUTH STATE COLLEGE

PLYMOUTH, NEW HAMPSHIRE 03264

Natural Science

October 11, 1967

OCT 13 1967

*mailing
list*

Dr. David Rogers
Dept. of Biology
Armory 101
University of Colorado
Boulder, Colorado

Dear Dr. Rogers:

I was able to run your program with my data before leaving for New Hampshire (one day before as a matter of fact) and have the printout sitting at my side although it may be a while before I get to interpret it as I am terribly swamped preparing for my classes. So far the fall has been gorgeous, but none-the-less, we still miss the mountains of Colorado. Thus unless bad times fall our way, we will be back visiting this coming summer and at that time I will inform you of my progress.

Thank you again for your assistance. I would appreciate receiving any publications that you might publish in the future concerning numerical taxonomy.

Sincerely yours,

Larry Spencer

Larry Spencer*

*Formerly of Colorado State University where I worked under Dr. E.B. Reed on amphipods.

30 October 1967

Warren E. Grabau
Research Geologist
U. S. Army Engineer Waterways Experiment Station
P. O. Box 631
Vicksburg, Mississippi 39180

Dear Warren,

Thank you for the very interesting article on computer art. I read it with interest as did Robert Brill, our computer expert here. Aside from the superficial amusement pursuant on perusal, these pictures suggest a further heuristic application, viz. "pictures" of familiar mathematical transformations, c.f. page 8, bottom. Perhaps a better intuitive feel for what various mathematical transformations "look like" might be provided in this way - especially for those whose formal mathematical training is insufficient for accurately envisioning the subtler properties of common transformations.

I am also anxious to see the results of our idea for coding ecological data. It would seem that we have addressed what has been a difficult problem with an encouraging bit of ingenuity. I intend to mention our idea at the forthcoming numerical taxonomy meetings at Lawrence, Kansas.

I look forward to our next meeting.

Very truly yours,

George F. Estabrook

GFA:gm

9 October 1967

Mr. Lester Talkington
53 Hickory Hill Road
Teppan, New York 10983

Dear Lester:

We have supplied all the requisite information to SHARE for the IBM 7044. We started using the IBM 7094 but ran into such difficulty with the WDPG for the western universities that we abandoned the project.

Enclosed are copies of the "Input Deck Setup," a paper George Estabrook may have left with you at the Symposium on how to set up characters, and a paper on "Reading the Printout."

We have not the faintest idea of how it will do with nutrition data since at least two things are unknown to us: (1) the nature of the data, and (2) what he wants the data to tell him. Our program satisfies the assumptions for the clustering of plants and animals that indicates relationships between either specimens of taxa for the eventual determination by the use of biological principles, of taxa or classification.

One additional output of the program is not in the write up. We call this the "Nearest Distance Array." It is simply a listing of the ten closest objects in order to each of the objects in the program along with the associated similarity value. We find it of use with reticulately related objects or for objects that lack significant amounts of information.

We are now running the program at the Graduate School's computer, a C.D.C. 6400, here at the University of Colorado.

Give my best to your wife and how are the birds migrating? If we can be of any more help let us know.

Sincerely yours,

Henry S. Fleming

HSF:gm

THE UNIVERSITY OF WISCONSIN
MADISON 53706

OCT 4 1967

DEPARTMENT OF ZOOLOGY
BIRGE HALL

30 Sept. 1967

David Rogers
Taximetric Laboratory, Armory 101
Dept. of Biology
University of Colorado
Boulder, Colo. ~~9~~80302

Dear Dave:

I was very impressed with what you and your group have done -- both in terms of the taximetric ~~discussion~~ applications discussed at the symposium and the methods used (after reading the papers given to me by Estabrook).

I helped a friend ~~of~~ use the Bonner Clustering Program on the data from a nutrition survey, with very promising results. I'd like to see how your program performs on his next study. But I don't even know what computer the program was written for, etc. Please tell me about the general characteristics and capabilities of the program, how to prepare data for it, ~~transfer~~ and the matter of how to arrange for use of the program or the processing.

Sincerely yours,

*Lester
Talkington*

Lester Talkington
53 Hickory Hill Road
Tappan, N.Y. 10983

PS: Write me at home. I go back to IEM on January 1st. Please forgive the atrocious typing but I have no secretary.

27 September 1967

Dr. Arthur S. Boughey, Chairman
Dept. of Population and Environmental Biology
University of California
Irvine, California 92664

Dear Dr. Boughey:

Thanks for your letter of the 13th. I am pleased to hear of your activities at Irvine. I congratulate you that you now have David Goodall working with you. I met him first in Texas. Our group consists of Henry Fleming (ent.-taxonomy), two mathematicians, Estabrook and Christensen, and two programmers, Brill and Handley. Our major effort at the moment is work on a computerized biological information retrieval system funded by NSF. We are just into this work. Our efforts are to reflect various parts of the systematic whole, and in our shop we have attacked the kind of problems we have (as you can see described in the reprints that I am sending separately). We have done a job in clustering and in character analysis, both of these have programs that may interest you. The clustering program has been flow charted and documented, and now resides in the SHARE Library as program #3501. We are in the process of flow charting and documenting the character analysis program.

Have you heard of the informal get-together in Kansas in November for numerical taxonomists? If not, you might write to Bob Sokal for information.

Sincerely,

David J. Rogers
Professor of Biology

P.S. We have completely exhausted our reprints of the first actual taxonomic job ever done with the computer, namely: "Monographic Studies in Cassia (Leguminosae-Caesalpinioideae). II-A Taximetric Study of Section Apoucouite, by Irwin and Rogers. Perhaps you can get a reprint from Howard Irwin, at the New York Botanical Garden.



DIVISION OF BIOLOGICAL SCIENCES
DEPARTMENT OF POPULATION AND
ENVIRONMENTAL BIOLOGY
IRVINE, CALIFORNIA 92664

September 13, 1967

Dr. David J. Rogers
Professor of Biology
Taximetrics Laboratory
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

Keith Justice passed to me your letter of 1 August. For the last two years we have been working on an interdisciplinary pilot project investigating the possibility of the production of computerized identification keys. I discussed this very briefly with Aske Love when he was over here, I suppose it was nearly six months ago, and he said at the time that a numerical taxonomy group was being formed at Boulder.

Our group here is presently composed of myself as Principal Investigator of this pilot study, Keith Justice as an Assistant Investigator, and David Goodall who has been working with us this summer. Keith could not get to the AIBS meetings, David Goodall attended the taxometrics session while I continued with the concurrent ecology program.

We have not published very much as yet, because we are moving so fast and changing things so rapidly that it does not seem worthwhile. However for our own purposes we put together technical notes. The early issues of these are already out-of-date, but I am just preparing a new one on the organization of the data base for a computer key, and we will shortly also prepare a second note dealing with our revised search strategy. If you would like to have these, I would be very happy to send them to you. Both may eventually be published, because the data base arrangement and search strategy are now crystalizing, and we have to have a little more stability if we are to produce an effective system. Meanwhile I would much appreciate receiving any literature which you are able to send.

I caught an occasional glimpse of you at College Station, but we did not come close enough for a conversation as all too frequently happens at these mass meetings, valuable as they are in other respects.

Best Wishes.

Yours sincerely,

Arthur S. Boughey, Chairman

SEP 15 1967

BRIGHAM YOUNG UNIVERSITY
PROVO, UTAH

84601



ERNEST L. WILKINSON, PRESIDENT

DEPARTMENT OF BOTANY
320 Brimhall

13 September 1967

Dr. Don Rogers
Department of Botany
University of Colorado
Boulder, Colorado

Dear Dr. Rogers:

As you suggested following the Taximetrics meeting I am writing you for reprints on the systems you use with your present model. As I indicated to you at that time, I was very much impressed with Frank Hawksworth's paper on Arceuthobium for his approach could be used in my study on the Genus Eriogonum as the Genus Eriogonum contains over 200 species and has both species and generic problems. A taxometric study would prove exceedingly interesting and probably very helpful.

I hope by reading your reprints I can see a new approach to the study of the Genus Eriogonum which your numerical model will help solve.

If I have the time and opportunity to study the Genus Eriogonum from a taxometric viewpoint I will write again for further information.

I remain.

Sincerely yours,

James L. Reveal

James L. Reveal

*mailed
19 Sept
Use of Computers
Info Theory Model for Char. Anal.
Math Model in Graph Theory
Graph Theory Model
Gen Method of Taxonomic Descrip.*

6 September 1967

Dr. Lester Talkington
Department of Zoology
University of Wisconsin
Madison, Wisconsin

Dear Lester,

I have just read with interest your recent paper "A method of scaling for a mixed set of discrete and continuous variables" which we discussed briefly at the recent AIBS meetings. I agree that your choice of Euclidean distance for illustrating your discussion is a good one as it is perhaps the most universally understood of the techniques extant. In its purest form, Euclidean distance can only be computed within a Euclidean space, i.e. an orthogonal copy of the real line. This required that you generalized the notion of Euclidean distance somewhat at the very onset of your discussion (p. 149 bottom of right column) to include a meaningful treatment of qualitative characters of more than two states which do not admit a natural ordering. Given that Euclidean distance be generalized as you have done, your following discussions are interesting and do indeed strike an appropriate balance between mathematical rigor and biological applicability. But let me discuss, for interest's sake, other possible generalizations of Euclidean distance, *sensu strictu*, to include qualitative characters. As I understand it, your approach is to "replace" a qualitative character with many quantitative characters as follows: Replace

Study	Character mapping	descriptive state 1
		descriptive state 2
		⋮
		descriptive state k
Study	Character mapping	applicability of descriptive state 1
		1 = applicable
		0 = not applicable
Study	Character mapping	applicability of descriptive state 2
		1 = applicable
		2 = not applicable
		⋮
Study	Character mapping	etc. for descriptive state k

In this way the actual computation of the Euclidean distance may proceed in the traditional arithmetic manner.

Then you continue by arguing that d^2 is made up as a sum of "contributions" as follows:

$$d_{ij}^2 = \sum_{p=1}^n (\text{contribution of character } p \text{ for object pair } i, j)$$

where the contribution for character p , when p is qualitative, takes the form of a sum within the above sum as shown

$$d_{ijp}^2 = \sum_{s=1}^K x_{ips} - x_{jps}$$

where I have taken the liberty of subscripting p for the sake of my proposed generalization. I agree that conceiving of d^2 as being made up by a sum of contributions is a useful and productive way to think. I would like to suggest that this point of view be generalized as follows:

$$d_{ij}^2 = \sum_{p=1}^n \phi_p(x_{pi}, x_{pj})$$

where

$\phi_p(x_{pi}, x_{pj})$ is some function associated with character p , this function assigns to the object pair (i, j) a contribution to the "distance" based on the observed states of character p for object i (x_{pi}) and object j (x_{pj}). This functional composition is illustrated below:

character

x_{pi}

character

x_{pj}

With the data manipulating features of today's computing machines there is very little to be gained in restricting ϕ to the class of function whose values can be arithmetically computed. To turn again to the method of your published discussion, observe that ϕ_p would take the form:

$$\phi_p(x_{pi}, x_{pj}) = (x_{pi} - x_{pj})^2$$

and in the case of qualitative characters

$$\begin{aligned} \phi_p(x_{pi}, x_{pj}) &= \sum_{s=1}^K (x_{ips} - x_{jps})^2 \\ &= 2 \text{ if } x_{pi} \neq x_{pj} \\ &= 0 \text{ if } x_{pi} = x_{pj}. \end{aligned}$$

Clearly if we have the definition of ϕ_p open, in the sense that it will depend on biological considerations private to character p , we can choose

to define it as you have, or we can define it in any other way deemed appropriate. In this way qualitative contributions need not range from 0 to 2, but can be made to behave as we wish. Of course, the question of how they ought to behave has been left entirely open at this point. I have

I have included a preprint of our BioScience article which addresses itself to this question somewhat. I'm sure that you will agree that under some of these other generalizing schemes for qualitative characters the relevance of your later discussions will be influenced somewhat.

Please be encouraged to reply to my comments if you feel that they are interesting to you.

Very truly yours,

George F. Estabrook

GFE:gm

Signed in his absence/

AUG 29 1967

UNIVERSITY OF COLORADO MUSEUM
BOULDER, COLORADO 80302

File

Dear Dave:

I spoke to Dean Archer about the Smithsonian business, but did not give him any names. He will talk to you and get names of contacts when and as you feel he will be ready to talk to them. He thinks we have a good idea, and this should go a long way to seeing a project through. Sorry I don't have time to see you before I go. Cheers!

Bill

William A. Weber
curator

Jim:

~~Re. Bill's proposal + mine, I think that it would be well to talk to Dr. Poulton W. Adkisson, head, Off. Sci. Info. Serv., N.S.F. Also Dr. Sidney Galter, Asst. Secy. for Science, Smithsonian Inst.~~

FILE COPY

UNIVERSITY OF COLORADO

BOULDER, COLORADO 80302

Taximetrics Laboratory
DEPARTMENT OF BIOLOGY
Armory 101

24 August 1967

Dr. Walter H. Bernotat
Eidg. Technische Hochschule
Institut für Kristallographie und Petrographie
Sonneggstrasse 5
8006 Zürich, Schweiz

Dear Dr. Bernotat:

I have to respond to your letter to Mr. Estabrook requesting a FORTRAN listing for our program for biological classification. I trust that your Institute possesses or rents equipment from IBM. If you do, it is possible to request from the SHARE Library listings and documentation and flow charts of this program. The address of this organization is: Program Information Department
International Business Machines
40 Saw Mill River Road
Hawthorne, New York 10532, U.S.A.

and the number of our program in the SHARE Library is: 3501.

If you have no connections with IBM (International Business Machines) and there are no IBM field representatives in Zurich who might make the request for you, we will be pleased to make available a listing from this office.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

RETYPE AND MAILED VIA AEROGRAMME

AUG 22 1967



Eidg. Technische Hochschule
Institut für
Kristallographie und Petrographie

Vorsteher: Prof. Dr. F. Laves
Telefon 32 73 30

Walter H. Bernotat

8006 Zürich, 16.8.67
Sonneggstrasse 5

Switzerland

Dear Dr. Estabrook,

I have read with great interest your papers on "A Mathematical Model in Graph Theory for Biological Classification" and "A Graph Theory Model for Systematic Biology". It seems to me that this method of classification is also applicable to rocks and rock series.

I would be most interested to obtain from you a FORTRAN-listing of your program (if possible with a calculated example) and also any other paper you may have published on this subject.

Thanking you in advance for your assistance in this matter,
I remain

very sincerely yours

Walter H. Bernotat

Taximetrics Laboratory

Armory 101

4 August 1967

Mr. Gary Godsey
240 W. Prospect
Fort Collins, Colorado 80521

Dear Gary,

I am enclosing a copy of the write-up for the Similarity
Clustering Program together with instructions for reading the
Printout. These are only slightly out of date. I hope you
don't have too much trouble with them.

Sincerely,

Robert C. Brill

RCB:gm

AUG 2 1967

DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTER SCIENCES

SCHOOL OF ENGINEERING AND APPLIED SCIENCE



WASHINGTON UNIVERSITY

SAINT LOUIS, MISSOURI 63130

August 1, 1967

Dr. David J. Rogers, Professor of Botany
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

Your letter to Dr. Burch came at a time at which I was heavily committed elsewhere. I am sorry I did not answer you before and I am sure that you will understand.

Certainly I will be very pleased to receive flow chart and documentation of your clustering program and make some arrangements to use this program for class (and, of course, also other) purposes. In addition I would be very pleased to receive some information on your work since this is an area of extreme importance and great interest to me. In fact, you might find our taxonomical approach quite useful. (I include one reprint on this work and will be happy to supply you with additional information if so desired).

However, it is the purpose of our course to give the student experience in handling data not only in specific heuristic clustering techniques but also in more mundane analyses. We prepare for the course pools of data describing varying types of biological phenomena. Dr. Burch thought that the data on which you base your work would be extremely useful for these purposes. Perhaps to explain this matter better it might be useful for you to read the enclosed reprint describing the general approach to our course. Please bear in mind that the work reported was for a course of thirty-two lecture hours while now we are teaching a two-semester course in mathematics of computation, numerical techniques and statistics useful for biological research workers.

What we would really like, thus, are data relevant to the fields of botany since a large part of the class will come from this department.

Sincerely yours,

Theodor D. Sterling
Theodor D. Sterling
Professor

TDS:hk

Enclosures

Signed in the absence of Dr. Sterling by Hilda Ketterer, Secretary

Simple Copy

- Taximetrics Laboratory

June 23, 1967

Dr. Derek Burch
Missouri Botanical Garden
2315 Tower Grove Avenue
St. Louis, Missouri 63110

Dear Dr. Burch:

I will be glad to correspond with Dr. Theodore D. Sterling concerning the possibilities of biological examples for courses tailored for biologists. I would like, however, to correspond directly with Dr. Sterling in as much as the data we have punched on the cards is relatively meaningless unless one knows the structure of the raw data, its intended purpose, etc. I would not think it would make a meaningful type of example for him. I have several suggested alternatives.

1. That we supply a flow chart and documentation of our clustering program which is well documented.
2. That we send also a deck of data which can be run through this program.
3. That we send a completed run so that Dr. Sterling can select from this much more meaningful problems and their solutions for mathematics.
4. That we send the mimeographed notes of the course which we offer dealing with the mathematical aspects of classification.

If this is satisfactory, please ask Dr. Sterling to communicate with me at my new address after July 1, 1967:

Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Sincerely yours,

David J. Rogers
Professor of Botany

MISSOURI BOTANICAL GARDEN
"SHAW'S GARDEN"



2315 TOWER GROVE AVENUE
ST. LOUIS, MISSOURI 63110
TOWNSEND 3-0440

June 19, 1967

JUN 22 1967

Dr. David J. Rogers
Department of Botany
Taximetrics Laboratory
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

I am presuming on two brief meetings at symposia here to ask for your help. We are starting a three-semester series of math courses tailored for biologists and need raw data from which to draw examples that have some meaning. Dr. Theodore D. Sterling of the Applied Math Department is teaching the course and will be setting up the problems this summer. He needs data which are already on card or tape but apart from this has no stipulations. He has assured us that the material would be used only for classwork and would probably be glad to give you the same guarantee in writing. If you think you might have anything that we could use I will ask Dr. Sterling to contact you directly. (!)

All is well at the Garden as we go into the heat of summer. Walter says that he saw you at Ann Arbor and that you may be in later. We will look forward to seeing you then.

Sincerely,

Derek Burch

Derek Burch
Assistant Botanist

Assistant Professor of Botany
Washington University

DB/dp

cc: Dr. Theodore D. Sterling

Taximetrics Laboratory

Armory 101

1 August 1967

Dr. Keith Justice
Department of Population and Environmental Biology
University of California
Irvine, California 92664

Dear Dr. Justice:

Dr. Peter Rauch was kind enough to indicate in a recent letter that you are interested in the development of models and computer programs for plant identification. We are much interested in this field and wonder if you have any papers describing your work. I am sending a set of reprints from our lab which will indicate the nature of our interests so far.

We look forward to corresponding with you. Incidentally we will be presenting a symposium at the AIBS describing some of our projects. It will be on Tuesday afternoon, 29 August, and is jointly sponsored by the Biometrics Society and the American Society of Plant Taxonomists.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

DATAMETRICS, INC.

P. O. BOX 318

BOULDER, COLORADO 80302

AUG 7 1967

Services in data processing and scientific data analysis

July 31, 1967

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado

Dear Dr. Rogers:

In response to your advertisement in the July issue of Data Processing Magazine, we at Datametrics, Inc., are offering our services in large scale computer data analysis and programming.

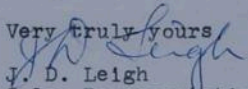
Because of the apparent parallel in the nature of your work and in our programming experience and capabilities, we feel that we are uniquely qualified to assist you.

We offer past experience utilizing both large and small machines in scientific work, and currently, we are able to offer proficiency in the use of the Control Data Corp. 6400 computer.

Datametrics, Inc., is a Colorado corporation located conveniently in downtown Boulder at 1117 Pearl, office #5A. Our telephone number is 447-1348.

We invite you to call us for information or further discussion. We will be pleased to serve you.

Very truly yours


J. D. Leigh
Sales Representative

Taximetrics Laboratory

Armory 101

31 July 1967

Dr. Peter A. Rauch
Department Of Entomology
University of California
Riverside, California 92502

Dear Dr. Rauch:

Thank you for your interesting letter of the 24th. I am sending along the reprint requested but am also sending other material of interest in connection with the graph theory model. In particular I am sending a mimeographed copy of a paper which appeared in Bioscience 1966. We are out of reprints but the copy contains all of the elements as well as directions for structuring the input data. The clustering program is quite sensitive and requires the biologist to be pretty much on his toes when he goes about the process of making up his characters. The instructions for structuring the input data follow as closely as possible the "normal" way of putting taxonomic data together, but still the biologist must pay close attention to this paper for best results with the graph clustering model.

I trust that your computer center at Riverside uses IBM equipment. If you do, you may request them to send to the SHARE Library for Program #3501. Included with the SHARE material should be a statement on "How to Read the Printout." If you do not have that piece of paper from SHARE, let me know, because the output from the material, which is easy to understand, still looks like gibberish unless you know how to use it. This has been written down, and hopefully will come to you with the package from SHARE.

If your center has no IBM equipment then write to us and we will have to fix you up with the program deck (or listing) and flow charts. Again, do let me know if you need these.

It is a pleasure to see others write down their opinion about the right way to use computing machines for taxonomic problems. We are certainly agreed and my whole fight is to do a job for taxonomy, not taximetrics or numerical taxonomy or quantitative taxonomy or any other set of names.

Thanks for the lead on Keith Justice concerning identification programs. I shall contact him right away.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

JUL 27 1967

UNIVERSITY OF CALIFORNIA, RIVERSIDE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

DIVISION OF AGRICULTURAL SCIENCES
COLLEGE OF AGRICULTURE
CITRUS RESEARCH CENTER AND
AGRICULTURAL EXPERIMENT STATION

DEPARTMENT OF ENTOMOLOGY
RIVERSIDE, CALIFORNIA 92502

July 24, 1967

AIR MAIL

Dr. David J. Rogers
Department of Botany &
Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

I would very much appreciate receiving a reprint of the article "Use of Computers in Studies of Taxonomy and Evolution" in Evolutionary Biology, Vol. 1, 1967, if available.

Also, if I may have a copy of the program for the graph theory model used in the orchid and/or the Cassia studies, I would find it very useful.

I am attempting to develop a computer-oriented logical system for aiding in the solution of taxonomic problems. My general philosophy towards this problem has been that very acceptable results are produced by many classical taxonomists who are specialists in particular groups. An attempt to simulate their approach seems more reasonable than any attempt to redefine the basic tenets and goals of classical taxonomy (a la Sokal, perhaps). Your comments on the "intuitive" approach, ^{and} on the "integrating system" used by practicing taxonomists reflect this observation also.

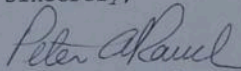
I once jokingly made the comment to a friend that the best model to develop would be a structureless one, a model that did absolutely nothing to the data. This came as a result of long discussions on the adverse effects produced by such operators as the distance function, which you point out, too. Perhaps my structureless model won't make you laugh as hard as he did.

Incidentally, you may be interested in contacting Dr. Keith Justice, Department of Population and Environmental Biology, University of California, Irvine, California, about his group's

Dr. David J. Rogers
July 24, 1967
Page 2

work to develop computer programs for plant identifications
(your "area IC", page 179-180). I don't believe they have
published yet as the task is still in the intermediate stages.

Sincerely,

A handwritten signature in dark ink, appearing to read "Peter A. Rauch". The signature is fluid and cursive, with the first name "Peter" and last name "Rauch" clearly distinguishable.

Peter A. Rauch

PAR:ab

Taximetrics Laboratory

Armory 101

11 August 1967

Dr. George Van Dyne
Department of Fishery and Wildlife Biology
242 Forestry Building
Colorado State University
Fort Collins, Colorado

Dear Dr. Van Dyne:

I trust you found a binary deck and source program listing for the Similarity/Clustering Program on your return from your trip. I am enclosing two documents which should be useful to you for setting up data cards for the program and for interpreting the results.

Should you wish to consult us further on the use of the program, please feel free to call on us.

Sincerely,

R. C. Brill
Programmer

RCB:gm

Taximetrics Laboratory

Armory 101

4 August 1967

Dr. Theodore J. Crovello
Department of Entomology
University of Kansas, Lawrence, Kan.

Dear Ted,

I have been meaning to reply to your note. It was a real pleasure to meet you at Ann Arbor. Before that it would seem to me that we were just names in sometimes not altogether friendly letters. I particularly enjoyed some of our less formal discussions with Steve Farris and Jim Rohlf.

I am sending along the reprints you requested. I might also draw your attention to a recent Irwin ^{paper} which appeared in the May 1967 Memoirs of the New York Botanical Garden.

Although I expect to be very busy this fall, I would most certainly be interested in a gathering somewhere if it can be worked into the schedule.

Best of luck with your new Notre Dame position.

Very truly yours,

George F. Estabrook

UNIVERSITY OF CALIFORNIA, IRVINE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

DIVISION OF BIOLOGICAL SCIENCES
DEPARTMENT OF POPULATION AND
ENVIRONMENTAL BIOLOGY
IRVINE, CALIFORNIA 92664

Dr. David J. Rogers,
Department of Botany and Plant Pathology
Colorado State University
FORT COLLINS
Colorado 80521

July 8th, 1967

Dear "r. Rogers,

I have in front of me your two letters dated April 17 - one forwarded to me from Western Australia, the other received here. I am sorry not to have responded to them earlier, but they were asking for separates of two recent papers, and it is only during the past two or three weeks that I have been reunited with all my papers, after their slow transit on a freighter across the Pacific. Anyway, I now have much pleasure in enclosing them, along with one or two others that I think may interest you, and I shall see that your name is on my mailing list for papers in this field in future.

note change

Best wishes !

Yours sincerely

David W. Goodall
Professor of Biological Science

Taximetrics Laboratory

Armory 101

19 July 1967

Dr. Lennart A. Bergstrom
Visiting Scholar - Urban Data Center
121 More Hall
Department of Civil Engineering
University of Washington
Seattle, Washington 98105

Dear Dr. Bergstrom:

I am sending herewith the papers you have requested. I am also taking the liberty of sending some further work of ours which has more definite value (in our estimation) for purposes of classification. The graph theory clustering program (mentioned in the reprints) has been deposited in the IBM Share Library - the program was written for the 7044 (it has also been rewritten for the CDC 6400). It is UQ CTCT Similarity Clustering Program of the Taximetrics Laboratory, Share Distribution No. 3501 in the Share Library. The model described in the Science article is not in the Share Library, but may be available as the IBM Taxonomy Model by Tanimoto and Loomis from the Math and Applications Department, I.B.M., Post Road, White Plains, New York.

Please note our new address.

Sincerely yours,

David J. Rogers
Professor of Biology

DJR:gm

UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98105

July 3, 1967

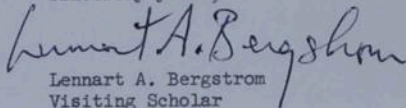
Department of Civil Engineering
Urban Data Center
121 More Hall

Mr. David J. Rogers
New York Botanical Garden
New York, New York

Dear Mr. Rogers:

In two articles in Science, 1960 and 1963, you discussed a computer program for classifying plants. This is a very interesting work and I would like to hear if you have a reprint of these two articles and if you have a write-up of your program or perhaps even a manual. Are you going to submit your work to the IBM Share Library?

Sincerely yours,



Lennart A. Bergstrom
Visiting Scholar

LAB:rjl

- Taximetrics Laboratory

June 26, 1967

Mrs. Judith Prewitt
Department of Radiology
University of Pennsylvania
3400 Spruce Street
Philadelphia, Pa. 10104

Dear Mrs. Prewitt:

Enclosed are the listing, the write-up, and instructions for reading the printout for the similarity-clustering program you requested. If you have any questions in the use of these materials, please feel free to write us.

Sincerely yours,

R. C. Brill
Programmer

RCB:ch

P.S. Please note our new address given on the first page of the write-up.

- Taximetrics Laboratory

May 29, 1967

Dr. Charles L. Remington
Department of Biology
Osborn Memorial Laboratories
Yale University
New Haven, Connecticut 06520

Dear Dr. Remington:

Thank you for your reprint request for our paper in Evolutionary Biology. I would be very interested in any "feed-back" which you or any students might care to make concerning this paper. The reason for this request is that we are gathering ideas for continued application of computers in evolutionary studies, and frankly need all the input we can get from students of evolution to make the most reasonable types of computer programs. Any comments or suggestions you care to make will be appreciated.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

Also included
P.S. Dr. Rogers' address after July 1, 1967

Yale University *New Haven, Connecticut 06520*

DEPARTMENT OF BIOLOGY

Osborn Memorial Laboratories

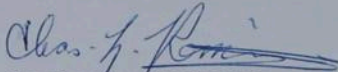
24 May 1967

Dr. David J. Rogers
Taxonomy Laboratory
Department of Botany &
Plant Pathology
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers:

If you have reprints available, I would appreciate your sending me a copy of your outstanding paper on "Use of computers in studies of taxonomy and evolution" in the new Evolutionary Biology. Having a paper in the next volume, I am aware that reprints are precious few in this series, but I hope you can spare one for my files. The latter get heavy use from a sizable group of evolutionists here in addition to myself.

Yours sincerely,



Charles L. Remington
Associate Professor of Biology

CR:rc

May 19, 1967

Dr. Jesse H. Shera
Dean, School of Library Science
Western Reserve University
Cleveland, Ohio

Dear Dr. Shera:

I had earlier planned to write to your school but I am glad I came across your paper "Librarians against Machines" in Science before doing so. My purpose in writing is to ask if any graduates from your school would be interested in working with me in a new information retrieval system beginning this summer. I was not able to write earlier because I have just been notified by the Office of Scientific Information Services of the National Science Foundation that our request for funds had been approved.

Very frankly, we are entering the field of information retrieval on what might seem to be amateur's standing. This is probably true in some sense but we feel some competence in that we have done work with computers and classification with some success for the biological sciences. Our approach is perhaps a bit different from the librarian's approach to information retrieval, but we feel a sufficient relationship exists such that we should definitely have working with us a person trained from the librarian's point of view. Our efforts will be experimental--not quite up to the development stage, at least initially. We would like to have working with us a young man or woman interested in the software aspects for the development of a generalized information retrieval system for biology. This team now has two biologists (myself and Henry Fleming), a mathematician and two programmers. We would like to have the expertise of a young and vigorous person trained with you.

If you have some person available, we can offer a position as a research associate at the University of Colorado, Boulder where we will begin this project July 1. We can offer up to \$9,000 per year for a 12-month appointment with a 10% increment the second year. While the grant which supports this effort is for only two years, we anticipate that the work will continue much longer and hopefully that there will be an opportunity for long-term employment at the University.

Dr. Jesse H. Shera

- 2 -

May 19, 1967

I will be glad to have your recommendations if you have someone willing and able to join our team.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

- Taximetries Laboratory

May 12, 1967

Dr. John Cutbill
Department of Geology
Sedgwick Museum
Downing Street
Cambridge, England

Dear John:

I think that we are close enough to the official notification of our grant for the information retrieval project to begin to think about the actual work to be done. As I recall, you had hoped to have someone to work with us during the development of the project. I hope this is still part of your thinking.

Since we saw you we have decided to move our operations from this university to the University of Colorado at Boulder, Colorado where our address will be: Department of Biology, University of Colorado, Boulder, Colorado 80302. There will be no shifts in the operation and we think that we can get ahead faster there than here.

I trust that we can look forward to some sort of collaboration and hope that you will be able to send someone interested in developing the necessary ties between us. Let me hear from you soon.

Sincerely,

David J. Rogers
Professor of Botany

DJR/ch

P.S. Are you aware of the IR meeting in Mexico City December 18-20? We (Brill and I) expect to be there. The meeting is jointly sponsored by the National Museum of Mexico and the Smithsonian, and focuses on the IR problems of large museums (musea?).

Received ^{sent} Lickig and flow chart for similarity-clustering
program

Mr. Thomas Bauer

5/19/67

U.S. Army Medical Research & Nutrition Lab.

Fitzsimons General Hospital

Denver, Colorado 80240

Division Chief

Col. James C. Syner

May 22, 1967

Mr. Richard Barnes
Associates for International Research, Inc.
Eight Eliot Street
Cambridge, Massachusetts 02138

Dear Mr. Barnes:

I am enclosing a sample of the output produced by the similarity-clustering program. In converting to the IBM 360, you should take note of the following problems.

(1) ^{MPDATA} ~~DATA~~ table (whose structure is described in the write-ups and in the listings) is unpacked and edited by a routine written in MAP (7044 assembly language). The editing function could easily be rewritten in FORTRAN. The unpacking function could be rewritten in FORTRAN too, but it would operate rather inefficiently. FORTRAN is not designed to manipulate information in less than full word modules. It can be tricked into so doing, sometimes quite effectively, but in this case, the performance would be a clumsy one. I would suggest rewriting the editing function in FORTRAN and rewriting the unpacking function in 360 assembly language. Watch out for the change from a 36-bit word to a 32-bit word. *word*

(2) All FORMAT statements for output assume a printer with 132 characters per line. If your 360 printer has fewer characters per line, you must reformat the output.

(3) The carriage control characters work on the 7044 as follows:

Carriage control character	Action before printing	Action after Printing
blank	none	1 space
0	1 space	1 space
1	eject page	1 space
+	none	none

If the 360 carriage control characters function differently, the carriage control character which begins each output FORMAT statement must be checked and in some cases changed. When working in this area, study the sample print-out enclosed for the desired print effect.

May 22, 1967

(4) In the 7044 operating system (IBSYS), the standard input tape is tape 5 and the standard output tape is tape 6. All READ 5 and WRITE 6 statements will have to be changed to fit the requirements of the 360 operating system.

(5) The statements CALL CHAIN(1), CALL CHAIN(2), etc., CALL CHNEXIT, are peculiar to the 7044 method for permitting program overlay (the facility for calling in and executing sections of the program one at a time). The 360 operating system may employ different conventions for this feature.

(6) All statements prefaced with \$ are directives to the 7044 operating system and will have to be replaced by their equivalents in the 360 system.

(7) Limits on table sizes will depend on the amount of core storage available to you. This may require changes in DIMENSION and COMMON statements and tests for overflow of tables.

I think that all statements not affected by the above considerations can stand as they are, but program conversion being the nasty job it is, you will in the course of your efforts probably discover other areas of incompatibility. I wish you success in your efforts and if I may be of further assistance, please contact me.

Sincerely,

R. C. Brill

RCE/ch

Enc.

ASSOCIATES FOR INTERNATIONAL RESEARCH, INC.

Eight Eliot Street, Cambridge, Massachusetts 02138
617-354-2133

April 11, 1967

Professor D. J. Rogers
Taxonomic Institute
Colorado State University
Fort Collins, Colorado

Dear Professor Rogers:

Thank you for responding so helpfully to the questions raised by Mr. Richard Barnes of our staff last Thursday. As a result of your suggestions, we have acquired a copy of the articles entitled:

- "Mathematical Models in Graph Theory for Biological Classification," by Estabrook;
- "Graph Theory Model for Systematic Biology," by Wirth Estabrook and Rogers; and
- "A General Method of Taxonomic Description....," by Estabrook and Rogers.

The articles are exceedingly interesting, and we feel they will be helpful to us in developing our own system for clustering and classifying social data. If it were possible, we would appreciate obtaining copies of the flow charts and listings for the programs described in your taxonomic papers. As Dick Barnes told you, some modification of your programs will be necessary to enable us to handle our somewhat larger data base. Possibly you would find it useful to have flow charts and listings of our amended "Rogers" programs after we have derived and debugged these.

Dick tells me that you stress the importance of conservatism and care in evaluating the result of any taxonomic classification. This, of course, is a problem whose magnitude is even greater than that which is found in developing a program. One of the reasons for your work's utility when applied to our needs is, we feel, that the problem of interpretation is somewhat lessened by your approach.

With our thanks,

Sincerely yours,

Lawrence W. Barss

Lawrence W. Barss

LWB:cpa

*sent flow charts &
listings on April 20, 1967 (Rls)
also we are taking a copy of the
programs when possible.*

March 10, 1967

Mr. John Graham
The Computing Center
Post Office Box 5445
North Carolina State University
Raleigh, North Carolina 27607

Dear Mr. Graham:

I am enclosing a set of listings for the similarity-clustering program, together with a sample of the output it produces. In converting to the IBM 360, you should take note of the following problems.

(1) MDATA table (whose structure is described in the write-ups and in the listings) is unpacked and edited by a routine written in MAP (7044 assembly language). The editing function could easily be rewritten in FORTRAN. The unpacking function could be rewritten in FORTRAN too, but it would operate rather inefficiently. FORTRAN is not designed to manipulate information in less than full word modules. It can be tricked into so doing, sometimes quite effectively, but in this case, the performance would be a clumsy one. I would suggest rewriting the editing function in FORTRAN and rewriting the unpacking function in 360 assembly language. Watch out for the change from a 36-bit word to a 32-bit word.

(2) All FORMAT statements for output assume a printer with 132 characters per line. If your 360 printer has fewer characters per line, you must reformat the output.

(3) The carriage control characters work on the 7044 as follows:

Carriage control character	Action before Printing	Action after Printing
blank	none	1 space
0	1 space	1 space
1	eject page	1 space
+	none	none

If the 360 carriage control characters function differently, the carriage control character which begins each output FORMAT statement must be checked and in some cases changed. When working in this area, study the sample print-out enclosed for the desired print effect.

March 10, 1967

(4) In the 7044 operating system (IBSYS), the standard input tape is tape 5 and the standard output tape is tape 6. All READ 5 and WRITE 6 statements will have to be changed to fit the requirements of the 360 operating system.

(5) The statements CALL CHAIN(1), CALL CHAIN(2), etc., CALL CHNKIT, are peculiar to the 7044 method for permitting program overlay (the facility for calling in and executing sections of the program one at a time). The 360 operating system may employ different conventions for this feature.

(6) All statements prefaced with \$ are directives to the 7044 operating system and will have to be replaced by their equivalents in the 360 system.

(7) Limits on table sizes will depend on the amount of core storage available to you. This may require changes in DIMENSION and COMMON statements and tests for overflow of tables.

I think that all statements not affected by the above considerations can stand as they are, but program conversion being the nasty job it is, you will in the course of your efforts probably discover other areas of incompatibility. I wish you success in your efforts and if I may be of further assistance, please contact me.

We have had numerous outside collaborators running studies through this program and we have been interested to discover the wide variety of classificatory problems to which this method has been applied. We would be quite interested to know what applications you have in mind. In particular, we are curious to know how the method might benefit your School of Design.

Sincerely,

R. C. Brill

RCB:ch

Enc.

NORTH CAROLINA STATE UNIVERSITY

THE COMPUTING CENTER
POST OFFICE BOX 5445
RALEIGH, NORTH CAROLINA
27607

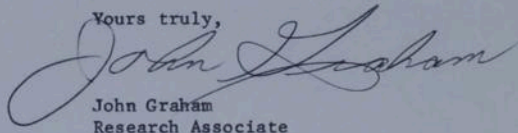
March 1, 1967

Mr. R. C. Brill
Taxonomy Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Mr. Brill:

Dr. Rogers has sent us the write-up for the similarity-clustering program. We have several people interested in it. One of these, a student in our School of Design, would like to convert it to our System/360. I think he will have less trouble if he had the listings in addition to the flow-charts. Could you furnish us with them? Also, do you know of anyone converting the MAP program to Assembler?

Thank you.

Yours truly,

John Graham
Research Associate

JG:gd1

- Taximetrics Laboratory

May 22, 1967

Dr. A. C. Smith
Department of Botany
University of Hawaii
Honolulu, Hawaii 96822

Dear Al:

You mentioned in your post card of May 9 that there are no students doing computer taxonomy there. Over the past year or so, I have been in correspondence with Miss Amy Jean Gilmartin who listed here address as Department of Botany, University of Hawaii. I note that after the 25th of February she is to be addressed in California. I was just wondering what her position is. She seems to be well grounded in the methodology.

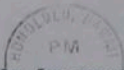
Sincerely,

David J. Rogers
Professor of Botany

DJR/ch

Dear Dave:

University of Hawaii, May 9, 67



ALWAYS USE

Many thanks for your prompt reply to my request and for the three reprints. All are excellent models to illustrate the use of computers and will be most useful to students. While the botanical taxonomists here are not doing this type of work, we like to know what is going on, and obviously the method has great possibilities.

Sincerely,

A. C. Smith





UNIVERSITY OF HAWAII • HONOLULU, HAWAII 96822

DEPARTMENT OF BOTANY, 2480 CAMPUS ROAD

May 1, 1967

5/4/67

Dr. David J. Rogers
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado

Dear Dave:

I have enjoyed reading your article (with collaborators) on the use of computers in the recent *Evolutionary Biology*. You have certainly provided a valuable commentary on the numerical method.

In a course I am giving on angiosperm evolution, some introductory time is spent discussing various methodologies, primarily as a means of looking into the validity of the phylogenetic philosophy. Your paper would be an excellent discussion for student review in a seminar.

In case reprints are available from this new journal, I would greatly appreciate having one at your convenience, so that my personal copy need not get hard student usage.

With thanks and best regards,

Sincerely yours,

A handwritten signature in dark ink, appearing to be "A.C. Smith".

A. C. Smith
Wilder Professor of Botany

ACS:dt

Taximetries Lab.

Mar. 15, 1967

Miss Amy Jean Gilmartin
Dept. Botany
Univ. of Hawaii
Honolulu

Dear Miss Gilmartin:

I am sorry to have delayed so long in answering yours of 2 Feb. Chalk it up to advancing age and poor organization. And I'm not going to answer all the questions you've asked at this time.

The one that seems to be most important to answer is your request that I read your paper. I am pleased to do so, at your convenience.

One easy question you asked: why the negative in the UN equation. That's merely to convert to positive numbers. Logs of numbers between zero and 1 are negative.

More later--sincerely,

David J. Rogers

Pantans. 3/16

2 February 1967

Dr. David J. Rogers
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers,

I owe you an apology. When I wrote what I thought was a follow-up letter having received the reprint which you so kindly sent, I was rather curious why you did not acknowledge my earlier letter. I see now that you had never received it. It was just returned to me having been misaddressed. I am sorry.

I enclose my earlier letter that ^{never} reached you. However, I see now that you have abandoned the earlier clustering method based on information theory. Fortunately this does not affect me too much as I was not doing anything with the clustering portion of the program, since there was something wrong with the values for inhomogeneity that I was obtaining. I do like the idea of that clustering method though and would still be interested in your comments in regards to question 2 in particular just for future reference.*

For my thesis which is an alpha-numerical taxonomic study of the Bromeliaceae of Ecuador I am using individual specimens for my OTUs. I will not have time to score more than approximately 200 specimens although I will be working with (preparing keys and descriptions) ~~with~~ several times that number.

I am more interested in the distance between OTUs and the variance of this distance with different randomly chosen set of characters using both the same number of characters and different numbers of characters than I am interested in some way of clustering or forming phenograms. It does seem to me that one very good point of attack for improvement in taxonomy through NT (in addition to your point of defining the activities of the taxonomists) is to define acceptable affinities, prerequisite for inclusion of two specimens in the same taxa. If we could put probability levels on a given questionable specimen belong in a given species and perhaps at the same time the probability of it belonging in a different species circumscription, it would be a big help to the alpha taxonomist. There certainly still is a great deal of alpha taxonomy to be done i.e. most of the tropical plants of the world. Anyway I think that this is an interesting problem which I would like to work on some day. However I won't get that far during my thesis research. I'm still at the point of trying to obtain a measure of the variance that accrues to my distance values as different sets (randomly chosen) of characters are applied. This is so that I can apply confidence intervals to all my distance values. Of course the variance will increase as the number of characters decreases from the maximum possible, 104 in my case. I am right now, trying with the help of a fellow from the mathematics department here to find an acceptable way to correct for this imposed increased variance which is due to having to choose subsets of characters from the same overall set. So far I can find no way better than the rough approximation of taking the potential difference for different ~~characters~~ ^{sets} and equating this to the maximum which occurs when 52 are selected. A solution would be particular useful to me since, as mentioned in the earlier letter, I have programmed so as to be able to use OTUs with any amount of missing data. This means, that I am using different sets of characters. This sort of difference has been shown to produce uncomfortable results. Have you seen Ted Crovella's PhD thesis (Berkeley, 1966)? He worked with Salix and had large numbers of characters applicable to one or the other sex. However, I am so far rather pleased with results from using different characters sets. For example,

* also most interested in your answer to question 4)!

resulting from
 with between 62 and 68 character comparisons, ~~chosen at~~ random ~~by~~ selecting
 of 280 characters, and a ^{mean} distance between two OTUs of 1.4024 ($n = 10$), the variance
 (s^2) was .004. With from 22-26 characters, applied to the same OTU pair, the
 mean distance was 1.388 and the variance was .092. If I crudely equate the
 necessarily smaller variance of the larger character number to the variance
 of the smaller characters number by multiplying the mean square of the smaller
 variance by 4 (since 80 out of 100 permits a maximum of 20 different characters
 between any two sets and since 50 out of 100 or fewer permits 100% possible
 difference between sets) I find that there is still a significant difference in vari-
 ance. The F test for heterogeneity of variance shows a significant difference at
 $P = .05$. However, it seems that there should be some more elegant method other
 than my simple muddled approach to accurately equate variances. So far I have not
 found any way nor has the math professor been able to do so. Any ideas?

As you can see I am really trying to kill ~~two~~ two birds with one stone:
 1) obtain confidence intervals for distance values, which confidence intervals
 will of course increase as characters number decreases and,
 2) find a minimum number of characters, a smaller number than which I will not
 consider valid for obtaining distance values. That is, if after correcting
 the calculated variance and comparing the corrected variances by means of an
 F test, I obtain an indication that the two variances different statistically
 then I would not use the smaller number of comparisons for purposes of obtaining
 distance values between OTUs.

At the same time I hope to be laying the ground work for attacking the problem of
 placing probability levels on any individual being included within a given species
 circumscription.

I hope that I will have time to set up my characters in order to be able
 to submit some OTUs to your program. I appreciate very much your kind offer
 to run some tests there for me. I really do not know if I will have time for
 that until after I finish the thesis. I hope so though.

Thankyou again and hoping to hear from you, I remain

Amy Jean Gilmartin
 Department of Botany
 University of Hawaii
 Honolulu, Hawaii 96822

after 25 February 1967

401 San Bernabe
 Monterey, California
 93942

P. S. In spite of the address change toward the end of this month, I am
 continuing as a graduate student at the U. of H. The mailing address only
 changes.

16 December 1966

Dr. David J. Rogers
 Department of Biology
 University of Colorado
 Boulder, Colorado

Dear Dr. Rogers,

Since I received such a nice letter from you in answer to one of ^{mine} ~~my~~ over a year ago in which I asked some question re. NT and your paper in BioSci. 1964, I am again taking the liberty to ask some questions.

At this time I have written a program for distance functions following your's and Fleming's model. I have included in the same program with the distance matrix print-out and T values, the inhomogeneity for the group of OTUs, and the standard statistical parameters using the distances themselves as the variables. (I realize that this is perhaps of only academic interest since the distribution of the distances will certainly not follow a normal distribution if the OTUs involved represent subsets having greater within-set affinity than between-set affinity) Additionally I have programmed for the number of comparisons which occur between OTUs; for the characters relevancy and the number of matches and mismatches produced by each character. This latter varies considerably because I have programmed so as to be able to compare OTUs even when there are ^{some} ~~some~~ comparisons, however, the print-out of Number of comparisons lets me know at a glance upon how many comparisons each of my paired comparisons is based. Finally, I have programmed so as to have a print-out whenever the ratio of actual attribute matches to potential attribute matches for each character is over .80 which gives me an idea about the association of attributes.

Now for the Questions:

1) inhomogeneity: Following Rogers & Fleming (1964) I have endeavored to obtain the correct maximum entropy and the correct relative entropy by programming as follows (Fortran IV): first I zero in G, then where ALL equals the number of OTUs I find the maximum entropy by ENMAX equal to the $\log_2(ALL * (ALL - 1.0) / (2.0 - G))$. Since the computer (IBM 7040) 1401 deals in natural logarithms I have divided the above by the $\log_e(2.0)$ which procedure for conversion has been checked out. Next, in order to obtain the relative entropy, RELEN, I have done as follows: RELEN equals $RELEN - D(I, J) / \text{SUM} * \text{ALOG}(D(I, J) / \text{SUM})$ and again have converted to log to the base 2 as above. Finally, UN equals $(ENMAX - RELEN) / ENMAX$. My trouble is in the values of inhomogeneity that I obtain. They seem to be out by an order of magnitude and worse are not consistent. That is, if the OTUs represent one species, my UN value may be higher than when the OTUs represent both some of the same species and some different species, whereas the second group certainly is a more heterogeneous one. The values of UN that I obtain are all consistently 1.9 plus and less than 2.

2) re. UN, could you check to see if the following equation is the correct one for relative entropy, which is represented on page 22 (Rogers & Fleming, 1964)?

$$\frac{1}{2} \leq \frac{d_{ij}}{\sum_j d_{ij}} \left[\log \left(\frac{1}{2} \leq d_{ij} - \log d_{ij} \right) \right]$$

In your paper as you know, a negative sign is placed before the $\frac{1}{2}$ for that equation and yet no indication of taking the negative is shown in the four steps

in which this equation is explained verbally. Perhaps this is part of my trouble? As written in Fortran you can see that I have not included any negative sign. Incidentally, I do not have to use the $\frac{1}{2}$ since my sum (SUM) is calculated on only $\frac{1}{2}$ the matrix.

3) Re. the justification for converting the similarity index to a negative log. The statement is made, p. 19, that it is not difficult to transpose the degree of similarity to an equivalent measure of distance. But why is this justified? The reason that I ask particularly is because ~~that~~ I am in the midst of making some comparisons of the change in distance as the number of comparisons changes. Since the distances are logarithms, a difference in distance for more distantly related pairs will be much greater than a difference in distance for closely related pairs. I wish to normalize the distance values but do not know if this, in turn, is justified. Of course, I could just make my comparisons between the same pairs of OTUs as the number of comparisons changes, which, in fact is what I am currently doing. Nevertheless, I am still curious about the justification ~~of~~ the logarithmic transformation here.

4) Having bombarded you with questions I now wonder if you would be willing to act as a critical reviewer for the NT portion of my Ph.D thesis. As you might imagine I am nearly on my own here as far as NT goes although I have corresponded with Rohlf and Sokal and more recently with Ted Crovello. I am in the second year of research for the dissertation, "The Bromeliaceae of Ecuador, An Alpha-Numerical Taxonomic Study". As you might guess from the title my OTUs are specimens.

Thanking you, I remain

Sincerely,

Amy Jean Gilmartin

Amy Jean Gilmartin
Department of Botany
University of Hawaii
Honolulu, Hawaii 96822

- Taximetrics Laboratory

May 22, 1967

Dr. George Hepting
Southeastern Station
United States Forestry Service
Asheville, North Carolina

Dear Dr. Hepting:

I was pleased to receive your cordial letter and the information on INTREDIS. As you suggest, I will give it careful analysis, discover questions of interest and hope that we may have a fruitful exchange of information later.

It is fairly certain that our information retrieval system grant request will be funded, but until we do receive the funds, I should not be overly optimistic. It is our ambition, amongst other things, to discover from a number of different workers the types of requirements of their systems such that when we write up our computer programs they will have "building blocks" to give individual information system users a maximum freedom in choice of retrieval methods.

Sincerely,

David J. Rogers
Professor of Botany

DJR/ch

- Taximetrics Laboratory

May 4, 1967

Dr. George Hepting
Southeastern Station
United States Forestry Service
Asheville, North Carolina

Dear Dr. Hepting:

We have heard of your information retrieval system for forestry and wondered if there is any descriptive materials concerning your methodologies and the direction of your work. I would be pleased to receive any descriptive materials which you have available.

We hope to embark into a rather ambitious program and development of a computerized information retrieval system shortly and will be very pleased to have any ideas which you may have developed along these lines.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

UNITED STATES DEPARTMENT OF AGRICULTURE
FOREST SERVICE

FOREST DISEASE RESEARCH
P. O. Box 2570
ASHEVILLE, N. C.

1670(4600)

May 11, 1967

Professor David J. Rogers
Dept. of Botany & Plant Pathology
Taximetrics Laboratory
Colorado State University
Fort Collins, Colorado 80521

Dear Professor Rogers:

I am enclosing information on the development and operation of my INTREDIS literature retrieval system, as requested in your letter of May 4.

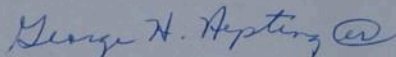
The first enclosure is our code book, the introduction of which gives a brief description of how the system works. Of course the development of the codes required considerable time and effort as you will see when you scan the book.

Also enclosed is a rough draft copy of my manuscript on INTREDIS which will be published soon in the "Agricultural Science Review." This will give you a clear picture of how the data are coded, then put directly into cards, then to tape, and finally computer scanned and printed out.

It would be impossible for me to explain in detail how INTREDIS was inaugurated and the many steps that were required to follow through, to become operational. As you can imagine, we have made various changes as we have gone along, in order to find cheaper and faster methods of getting the information to the public.

I might suggest that after you have studied the enclosures, that you get in touch with me again. By that time you may have some specific questions that I might be able to help you with.

Sincerely yours,



GEORGE H. HEPTING
Principal Research Scientist



Enclosures:
Code Book
Manuscript
Sample Punched Card

CODES TO THE INTREDIS REGISTER SYSTEM
FOR LITERATURE RETRIEVAL
IN FOREST PATHOLOGY

Compiled by
George H. Hepting
Forest Service
United States Department of Agriculture

Under the sponsorship of the
FAO-IUFRO Symposium on
Internationally Dangerous Forest Diseases
and Insects
Held at Oxford, England, July 20-29, 1964

AS 1000

INTERNATIONAL TREE DISEASE REGISTER

George H. Hepting

An electronic system of literature storage and retrieval for forest pathology has been put into operation by the U.S. Forest Service in cooperation with the Canada Department of Forestry and Rural Development.

Designed for worldwide coverage of the field of tree diseases, the International Tree Disease Register (INTREDIS) offers a much-needed service to researchers working in forest pathology. The project was sponsored by the FAO/IUFRO Forest-Pathology Symposium on *Internationally Dangerous Forest Diseases and Insects*, Oxford, 1964.

INTREDIS enables any qualified person to receive a literature list by author, approximate title of article, and reference to an abstract journal carrying the article. Coverage includes all articles likely to be of interest to a tree pathologist which have been abstracted by Review of Applied Mycology, Forestry Abstracts, Biological Abstracts, and Association Abstracts, Pollution Control, Abstracts of Air ^{will} Abstract citations dating from the early 1940's through 1965 have already been coded; ultimately, those dating back to 1930 ^{will} also be included.

INTREDIS can provide a printout within an hour after receipt of a request, and it is inexpensive. A request may be made by subject (for example, root rot or seed diseases), or kind of tree (for example, white oak or Douglas-fir), or causal organism (for example, Endothia parasitica or Pseudomonas syringae), or place (for example, Mexico or Australia or all of South America), or date, or author. Thus, the computer can search the magnetic tape in one operation for, say, "nematodes in the genus Xiphinema attacking white pine roots in Canada," and print out a list of appropriate references, as easily as it can print out the world literature on air pollution effects on trees.

✓ INTREDIS has some novel features that make it feasible, effective, and reasonably inexpensive. First, a professional forest pathologist ^{assigned to INTREDIS} selects the articles for coding by going through each issue of the named abstract journals. His knowledge of the discipline helps him determine what to select and how to code his material for the punch card operator, who punches one or more 80-column cards for each coded reference, the number depending upon the amount of cross-referencing needed.

✓ It is a cardinal point that the selector and coder be a professional forest pathologist. Who else would know to code the title "Dutch elm disease in winged elm" under host plant Ulmus alata and cause Ceratocystis ulmi; or that the "spike disease of sandal" ^{to be coded as} is a virus-caused disease of Santalum? Thus, we do not have incidents like the physiologist whose book "The Translocation of Food in Plants" came out under the subject of "belt conveyors."

The second unique feature of INTREDIS is that it is tied directly to the few abstract journals that cover well the field of tree pathology. Since our literature printouts show the volume and page number from these widely used journals, almost any scientist in our field will have ready access to the citations listed on a printout, and he will then get an abstract, as well, for each. By then, he should know if he wishes to see the full article or book.

The coder, using our published INTREDIS code book, writes out on a worksheet -- for each citation he selects -- the subject, host plant, causal organism, and country. All of these are in code. Then, using numbers and English letters, he enters the name of the abstracting journal, its volume and page, the year the article was published, the author, and an abbreviated title. A card punch operator transfers the worksheet information to coded cards; the card information is then put on magnetic tape.

A computer program for INTREDIS was written in a few days and put on tape also. An inexpensive computer system is used, since our operation is a simple scan (sort) and printout, with no computations to perform. At present our tape contains about ^{13,000} 10,100 citations. It can be "read" and a printout provided in about 15 minutes at a cost of about \$20 for machine time. ^{for whatever topic has been sensed}

I have been told that many schemes for electronic literature retrieval have failed because the quality of input was too low. In short, too often the selecting and coding was not done by people who really knew the literature in a given field. Other attempts have failed because what worked well for one field proved to be very poor for another. An economist who aided me in setting up INTREDIS said later: "This system works so well for forest pathology -- a well-abstracted field with good access terminology -- it is a pity I can't use it for my own field."

Some systems try to do too much and end up doing nothing. One system, expensive to subscribe to, gave me extremely little for a year's search because it tried to cover all sciences. As a result, there was little search of our particular source material. Systems based on key words in a title are cumbersome for what you get. Again they cover too wide a field, and anyone who has scanned some of these lists knows how long it takes to find an article he is interested in. There is another drawback to the key-word-in-title idea. If, for example, you wanted a particular article on air pollution injury to trees, you might have to look under six or more key words to find the article since the author may have used one of many alternate terms. He may have chosen to call it "SO₂ injury to trees," or "fume injury to trees," or "smoke injury to trees," or "air pollution injury to trees." With INTREDIS, however, you would get all air pollution articles by one simple sort on subject code 70. A quick scan of the printout would produce the specific articles you were looking for.

Finally, once set up, INTREDIS is easy to operate. When I get a request I tell my secretary what scans to make. She delivers the appropriate computer program cards to the computer center, and within a short time the citation printout is made and is on its way to the requester. Although we are still on a "pilot plant" basis, we have already handled requests from many countries and more than 50 individuals. Tapes are now being run both by the Canada Department of Forestry ^{and Rural Development} and the Forest Service of the U.S. Department of Agriculture.

Eventually, when our tapes are expanded to ^{cover} literature back to 1930, we also hope to be able to turn the operation over to the National Agricultural Library in Washington, D.C. A charge ^{may} ~~will~~ then be made for the machine time required to handle a request, but it will still be inexpensive. Even now, a 300-item printout on the world literature on Tomus annosus root rot -- that ordinarily would have taken months to compile -- was done by INTREDIS at a machinetime cost of \$20.

At present the service is offered without cost to United States and Canadian Government scientists and to others who have a pressing need for it. Requests for printouts should be addressed to the author at U.S. Forest Service, ^{P.O.} Box 2570, Asheville, N.C., 28802.

- Taximetrics Laboratory

May 8, 1967

Dr. M. L. Gay, Jr.
Dept. 72-14, Zone 401
Lockheed-Georgia Company
Marietta, Georgia 30061

Dear Dr. Gay:

I am sorry not to be able to provide you with very much information about XPOP other than what appears in the published literature. I am sure, however, that you can gain a considerable amount of information from the developer of XPOP, Mr. Mark Halpern, whose address is: Dept. 52-40, Bldg. 201, Lockheed Missiles and Space Co., Palo Alto, California 94304.

We intend to use XPOP as a general system on which we will build a generalized IR system. This has not yet been done. The TAXIR system will be but one output from the general system as an example of an application of the methodology for IR which we hope to make, given funds from NSF. Frankly, we are a little puzzled and would be pleased to know where you read about our methodology. The reason for quandary is that other than in a proposal to NSF we have not written down any discussions about the system. Since yours is the second inquiry, this is a little embarrassing, and I will appreciate hearing from you as to where the information was printed.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

LOCKHEED-GEORGIA COMPANY

A DIVISION OF LOCKHEED AIRCRAFT CORPORATION
MARIETTA, GEORGIA 30061



April 27, 1967

MAY 2 1967

Dr. David J. Rogers
Professor of Botany
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers:

I am engaged in information retrieval research with the Lockheed-Georgia Research Laboratory. Recently I read that you were using "a general compiler, developed by Lockheed Aircraft Company" in your Taxometric Information Retrieval System. Is this the "XPOP" compiler?

I am interested in the development of retrieval languages in general, and particularly in what use XPOP might be in this area. Any information on your system and the use of the Lockheed compiler, be it XPOP or not, would be appreciated.

Very truly yours,

LOCKHEED-GEORGIA COMPANY

M. L. Gay, Jr.
Assistant Scientist
Dept. 72-14, Zone 401

MLG:rw

- Taximetrics Laboratory

April 20, 1967

Dr. C. Gordon Bell
Department of Computer Science
College of Engineering and Science
Carnegie Institute of Technology
Schenley Park
Pittsburgh, Pennsylvania 15213

Dear Dr. Bell:

Other than the three references given below*, we have little to say about the compiler system developed by Mr. Mark Halpern whose address is: Department 52-40, Bldg. 201, Lockheed Missiles and Space Co., Palo Alto, California 94304. We have not begun the development of the TAXIR system and have only given the broad outlines of its features to the National Science Foundation in a grant proposal and to some people at the Smithsonian.

I am sending three papers to you from which you may be able to gather the information which you requested concerning "general articles on taxometrics." Perhaps these will give you the input necessary to discover our direction and aspirations, though incompletely. I have asked that my secretary place your name on our mailing list so that when anything does appear, you will receive copies. Since our grant has not yet been announced, it is not certain that we will be able to flesh out the bones of our ideas.

I am sorry that we cannot say anything definite yet but I think you can understand that we are somewhat reluctant to make any claims before we have done work ourselves.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

Enc.

* References attached

References:

- HALPERN, Mark I. Machine Independence: Its Technology and Economics. Research Laboratories, Lockheed Missiles & Space Company, Palo Alto California. 12 pages.
- HALPERN, Mark I. XPOP: A Meta-Language Without Metaphysics. Proceedings-- Fall Joint Computer Conference, 1964. pages 57-68.
- HALPERN, Mark I. Evolution of the Programming System. Datamation, July, 1964. Pages 51-53.

CARNEGIE INSTITUTE OF TECHNOLOGY
SCHENLEY PARK
PITTSBURGH, PENNSYLVANIA 15213

DEPARTMENT OF COMPUTER SCIENCE
COLLEGE OF ENGINEERING AND SCIENCE

TELEPHONE: 621-2600
AREA CODE 412

April 17, 1967

Dr. David J. Rogers
Professor of Botany
Colorado State University
Fort Collin, Colorado

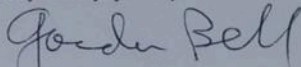
Dear Dr. Rogers:

I am very interested in receiving information on the compiler developed by Lockheed Aircraft Company which you have been using in your work on the Taxim System. Would you be kind enough to forward to me anything you have available in regard to it.

In addition, I would also be pleased to receive general articles on taxometrics.

Thank you.

Very truly yours,



C. Gordon Bell
Associate Professor of
Computer Science and
Electrical Engineering

CGB/dmj

UNIVERSITY OF CALIFORNIA, IRVINE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

DIVISION OF BIOLOGICAL SCIENCES
DEPARTMENT OF POPULATION AND
ENVIRONMENTAL BIOLOGY
IRVINE, CALIFORNIA 92664

Dr. David J. Rogers,
Botany Department
Colorado State University
FORT COLLINS
Colorado 80521

April 30th, 1967

Dear Mr. Rogers,

I have just received a separate of your paper,

"A general method of taxonomic description for a
computed similarity measure."

forwarded from Australia. I left there two months ago to take up a
new appointment here, and arrived less than a week ago.

Now that we are working within a thousand miles of one
another, I hope that we shall be able to make more effective contact,
for our interests are, in many respects, very close. Anyway, I
hope that you will continue to ^{send} me separates, and alter my address
in your mailing list to read as at the head of this letter.

Best wishes !

Yours sincerely

A handwritten signature in cursive script, reading "David W. Goodall".

David W. Goodall
Professor of Biological Sciences

- Taximetrics Laboratory

April 13, 1967

Dr. B. D. Ainscough
Department of Zoology
University of the West Indies
St. Augustine, Trinidad

Dear Brian:

I look back with great pleasure on my visit to Trinidad. It was indeed a pleasure to make your acquaintance and that of others who were so generous and friendly. Coming back to the frozen north country was no fun. I am sending under separate cover another copy of the flow charts for the similarity-clustering method along with several other papers which you may or may not have already.

I do hope that you will have a chance to visit with us at length sometime. Let me know if there is anything I can do to help out in that direction.

Once again, let me express my pleasure for your kind hospitality.

Cordially,

David J. Rogers
Professor of Botany

DJR/ch

Sent: Science 132(3434) 1960
BioScience 14(9) 1964
Brittonia 15(4) 1963
Systematic Zoology 15(1) 1966
J. Theoret. Biol. 12 (1966)
BioScience 16(11) 1966
General Write-up for Similarity-Clustering Program

UNIVERSITY OF THE WEST INDIES

CABLE: "STOMATA" PORT-OF-SPAIN
TELEPHONE: 662-5511



ZOOLOGY DEPARTMENT
ST. AUGUSTINE,
TRINIDAD, W.I.

19 April 1967

Dr. David J. Rogers,
Taximetrics Laboratory,
Dept. of Botany and Plant Pathology,
Colorado State University,
Fort Collins, Colo. 80521

Dear Dave,

Many thanks for your letter of April 13th. I must say how much I enjoyed meeting you and the various informal discussions we had. I look forward to happy relations in the future. The Tobago trip was very enjoyable, although the BWIA officer in Tobago was a bit suspicious of my return ticket. As the only "local" person on the trip I found myself appointed tour leader, so managed to bluff my way through the officialdom.

After leaving you I discovered that you had left your correspondence file in my car. Assuming you wish to keep it, I enclose it here.

I look forward to receiving the flow charts and reprints you are sending.

Very best regards,

Brian D. Ainscough,
Lecturer

Encl.

P.S.

I think you should start thinking seriously about publishing a successor to "Sokal & Smith".

B.

- Taximetrics Laboratory

February 23, 1967

Dr. Brian D. Ainscough
Zoology Department
University of the West Indies
St. Augustine, Trinidad, W.I.

Dear Dr. Ainscough:

Thank you kindly for your offer to help set up something for me on your computer.

It is not my intention to use the computer at all because I think that the participants will be more interested in the output than in the hardware. I will bring with me samples of the print-out and flow charts which should do the job. If all goes well, I will have a short 16 mm movie to illustrate some of the computer work.

Looking forward to seeing you,

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

UNIVERSITY OF THE WEST INDIES

CABLES "STOMATA" PORT-OF-SPAIN.
TELEPHONE: 662-5311



ZOOLOGY DEPARTMENT
ST. AUGUSTINE,
TRINIDAD W.I.

15 February 1967

Prof. D. J. Rogers,
Taxonomy Lab.,
Dept. of Botany and Plant Pathology,
Colorado State University,
Fort Collins, Colorado 80521

Dear Dr. Rogers,

As you will know, Mr. Haynes has asked me to organize any demonstrations, etc. that you may wish to set up in connection with your part in the Root Crop Symposium in April, due to my personal interest in numerical taxonomy, computer methods and also as I have already contacted you previously on these subjects.

The information you will wish to have concerning our facilities are as follows. The computer is an IBM 1620, with 20K core storage, Automatic Divide and Indirect Addressing facilities, and with paper tape input/output. It is possible that by the time of the symposium we shall also have card I/O, but this cannot be counted upon. In any case we do have a card-to-tape converter, so that any programs and/or data already on cards do not pose a problem.

I recollect that your previous studies have been done on a 32K machine, so that some alterations will obviously be necessary. Although these could be carried out when you arrive I would like to suggest that as many arrangements as possible in this regard be made in advance. Time will, no doubt, be rather shorter than one hopes for.

I should be most grateful, therefore, if you could let me know, at your early convenience, what arrangements you would like me to make. I do not anticipate any difficulties as regards securing computer time, either during the day or at night.

Yours sincerely,

A handwritten signature in dark ink, reading "Brian D. Ainscough".

Brian D. Ainscough,
Lecturer in Zoology

do not file yet

= Taxonomy Laboratory

September 23, 1966

Dr. Brian D. Ainscough
Zoology Department
University of the West Indies
St. Augustine, Trinidad, W.I. ==

Dear Dr. Ainscough:

I have to reply to your letter addressed to Dr. Michael Wirth at this address. Since Dr. Wirth is no longer with us, we will fill your request for a program listing. The program is now written in FORTRAN IV for the IBM 7044. When I have conferred with my programmer, I will determine the scope of the problem for the conversion to the 1620. I have no doubt that a clever programmer can transfer from the larger to the smaller machine, if you are willing to live with the time expenditure which the 1620 will require with this rather complex program.

The CDC 1604 mentioned in our paper used a language (FORTRAN 63) not compatible with the IBM's compiler system.

Since the program is extremely sensitive to the structure of the input data, I am taking the liberty of sending to you a pre-print which should be well digested by you before consideration of the program itself. This paper addresses itself to the problems of characters used in the taxonomic methodologies for computers and will, I am sure, be of value to you as you go into the taxonomy of soil mites.

Upon its completion, we will forward to you a program flow chart and listing. Thank you for your inquiry.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

sent flow chart & listing Nov. 28, 1966

UNIVERSITY OF THE WEST INDIES

CABLE: "STOMATA" PORT-OF-SPAIN
TELEPHONE: 862-5511



ZOOLOGY DEPARTMENT
ST. AUGUSTINE,
TRINIDAD, W.I.

17 September 1966

Dr. Michael Wirth,
Taxonomy Laboratory,
Dept. of Botany & Plant Pathology,
Colorado State University,
Fort Collins, Colorado 80521,
U.S.A.

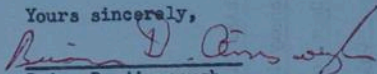
Dear Dr. Wirth,

I have read with interest your recent paper on the use of graph theory in numerical taxonomy (Syst. Zool. 15: 59-69).

I am engaged in a study of the systematics of a group of soil mites, mostly undescribed species and genera. At present I am investigating the possible use of numerical methods for the study, and would like to test the methods you have evolved. I would very much appreciate it, therefore, if you could send me a copy of your programs.

I am not familiar with the C.D.C. 1604 used in your study, however, and do not know what language it accepts or the format of your input (tape, card, etc.). The machine I am using here is the IBM 1620, with 20K memory and tape I/O (we hope to acquire card I/O in the near future), programs for which are written in either FORTRAN or SPS. I would very much like to know, therefore, whether your programs are likely to be compatible. In any case, a program write-up would be most useful.

Yours sincerely,


Brian D. Ainscough,
Lecturer in Zoology

UNIVERSITY OF COLORADO

BOULDER, COLORADO

GRADUATE SCHOOL

March 8, 1967

COMPUTING CENTER

Professor David J. Rogers
Taximetrics Laboratory
Botany Department
Colorado State University
Fort Collins, Colorado 80521

Dear Professor Rogers:

In response to your letter of March 2, we have checked the time on your job number 56073 dated 3-1-67. Our records indicate a running time on that job of 301 seconds. I assume that this figure agrees with the figure on your "tail sheet." It should be pointed out that our timing is a tape-to-tape time which is nearly independent of the quantity of printed output, representing primarily computation time in the CPU.

Perhaps you will remember sitting opposite me at a meeting of the Environmental Society a few weeks ago in Denver. I want to tell you again how much I enjoyed making your acquaintance at that time and how much I look forward to meeting you again in the future. If there are any further questions in connection with the job mentioned in your letter or in connection either with work we have already done or with work you would like us to do, please contact me at your convenience. I trust that I have answered your questions satisfactorily and repeat that I look forward to seeing you again.

Yours sincerely,



Ray Pitsker
Assistant to the Director

RP/led

- Taximetria Laboratory

March 2, 1967

Director of Computer Center
University of Colorado
Boulder, Colorado

Dear Sir:

Will you please check the accuracy of the time for job
#56073, date 03/01/67? The time seems inordinately long for
the output.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

- Taximetrics Laboratory

April 17, 1967

Dr. P.H.A. Sneath
Medical Research Council
Microbial Systematics Research Unit
University of Leicester
University Road
Leicester, England

Dear Peter:

May I please have copies of the following:

1. Mathematics and classification from Adanson to the present (Adanson Bicentennial Symposium, 1963, Hunt Botanical Library, Pittsburgh). Pp. 471-498 in Part II of Adanson, the Bicentennial of Michel Adanson's "Familles des plantes", edited by G.H.M. Lawrence. Hunt Botanical Library, Carnegie Institute of Technology, Pittsburgh, Pennsylvania. (1964)
2. Computers in bacterial and botanical taxonomy. Abstracts of Papers of the Xth International Botanical Congress, Edinburgh, 1964, p. 129. (1964)
3. The application of numerical taxonomy to medical problems. Pp. 81-90 in Mathematics and Computer Science in Biology and Medicine. Proceedings of conference held by Medical Research Council in association with the Health Departments: Oxford, July, 1965. London H.M.S.O. (1965)
4. Relations between chemical structure and biological activity in peptides. J. Theoret. Biol. 12, 157-195. (1966)
5. Estimating concordance between geographical trends. Systematic Zool. 15, 250-252.
6. A comparison of different clustering methods as applied to randomly-spaced points. Classification Soc. Bulletin 1 (No. 2) 2-18.

Thanks.

Sincerely yours,

David J. Rogers
Professor of Botany

- Taximetrics Laboratory

March 28, 1967

Dr. P.H.A. Sneath
Medical Research Council Microbial
Systematics Research Unit
University of Leicester
Leicester, England

Dear Peter:

I was very pleased to have your paper "Trend-surface analysis of transformation grids." I think you are onto something here. I would have used a different set of words to describe it but I don't understand the math either. I think one of the nicest things you said was what you referred to as a digression (page 102).

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

P.S. Our address will change July 1 from that given on the letter-head to: Taximetrics Laboratory, Department of Biology,
University of Colorado, Boulder, Colorado 80302.

M.R.C. Microbial Systematics Unit,
Leicester University,
Leicester, England.

13th December, 1966.

Dr.D.J.Rogers
Taximetrics Laboratory
Dept. of Botany and Plant Pathology
Colorado State University
Fort Collins
Colorado 80521
U.S.A.

Dear David,

Many thanks for your letter, and the enclosures. Michael Sackin and I have much enjoyed reading through the specification of your method and program. This is excellent for our newsletter and will be of great help to us. Yes, you have certainly sufficiently acknowledged our earlier work, and I am glad to see that you have carried it further. It would, I think, be very good if you could put a note in Taxometrics. I am in close touch with Hill, and I think you can take it that he would be glad to have about one page of quarto typing, which is usually about the limit of what he can take on any one topic. He has before now, when such matters have arisen, asked me to say that he much prefers the author to make the abstract, if he is willing, to avoid confusion over which parts are most appropriate. His address is:

L.R.Hill
National Collection of Type Cultures
Central Public Health Laboratory,
Colindale Avenue,
London N.W.9.

I found of great interest your paper to appear in Bioscience and hope to have a reprint of this fairly soon. I am sorry that Dr.Stearn has not acknowledged your work for him, but he was somewhat off-hand when some pilot work I did for him myself was sent to him!

I would like to hear more of Dr.Prance.

With best wishes for Christmas,

Yours sincerely,


P.H.A.Sneath

REGISTERED NO. 05006

Value \$ N.Y. Spec. del'y fee \$Fee \$ 75 Ret. receipt fee \$

Surcharge \$ Rest. del'y fee \$

Postage \$ 2.40 ☐ AirmailPostmaster, By David J. RogersFrom C.S.U.To P. H. A. Smith
Leicester University
Leicester England

FOD Form 2006-May 1964

4-6-10-72450-6



- Taximetrics Laboratory

November 16, 1966

Dr. P.H.A. Sneath
MRC Microbial Systematics
Research Unit
Leicester University
Leicester, England

Dear Peter:

At long last I am acceding to your request for documentation, flow charts, as well as a reprint of our graph theory clustering method. I am also enclosing a mimeographed copy of "Reading the Printout." With these documents you have as complete a set of instructions as I can put together. Since this documentation flow chart was done on an IBM 7044, it should be perfectly compatible with an IBM 7090 which I understand is available somewhere in the U.K. The flow chart is sufficiently well constructed and the documentation sufficiently complete, such that it should be easy to convert to other hardware.

I hope that we have sufficiently acknowledged your contributions. If not, please be assured that I recognize them. May I ask you about notification of this program to be made in Taxometrics or in some similar news bulletin for those interested in such methodologies. I am sure that Taxometrics does not wish to reproduce all that we have sent you. It might be appropriate, however, to take the comments for each of the subroutines, extract them and print those in Taxometrics. Would you be kind enough to suggest to Hill (or whoever is running Taxometrics now) that he take these comments and publish them? This should give interested workers enough insight to decide the applicability of this method.

Another document is also included. This is a preprint of a paper to appear shortly in BioScience entitled "A general method of taxonomic description for a computed similarity measure." Since the computer method is very sensitive to the data structuring, we have addressed ourselves to this problem in this paper. Anyone who attempts to use the method would be best served to consider the suggestions made for data structuring given in this paper.

We have tested the method on several different groups of organisms using data supplied to us by other workers. Two of the workers whose data we have tested are countrymen of yours. One you know quite well,

Dr. Stearn, of the British Museum. Unfortunately, Dr. Stearn has not so much as acknowledged the receipt of our printout so we do not know if he is dissatisfied with the outcome or not. The other is a young man, G. T. Prance, whose degree is from Oxford and who for the past several years has been working at the New York Botanical Garden. His study on the genera of the Chrysobalanaceae was a comparative one in which he employed several computerized methodologies for clustering. The results of his work are in the press in *The New Phytologist*. Other studies of this size group are in various stages of preparation for publication. Our effort has been to explore the value of the method in a rather broad spectrum, to discover where we may improve.

I trust that this methodology will be useful to you. I will welcome your comments and questions.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

Put this name and address
in our Rolodex address file

FEB 23 1966

MRC Microbial Systematics Research Unit,
Leicester University,
Leicester, England.

Tel. Leicester 50000.

Dear Colleague,

We will shortly be preparing a list of computer programs on classification and related topics, which will be circulated as an information service.

For this we would be glad if you would send us particulars of your programs on the attached form. Even if you have already sent us a description, it would be a convenience if you would return this form, since many programs are continually being revised.

Please also send full specifications if possible, both for our program files, and so that we can amplify the descriptions where required. Working documents and uncompleted write-ups will also be welcome.

Unless you have any objection, we would like to pass on brief details (based on your specifications) of completed programs for the next issue of Taxometrics.

P.H.A. Sneath

M.J. Sackin

MRC Microbial Systematics Research Unit,
Leicester University, Leicester, England.

Dr. Rogers.

We are just completing the list mentioned above, and owing to a regrettable oversight we think we have not sent you this circular. If you would like your programs to appear, would you be good enough to return the details Air Mail by return of post.
(CDC 1604 programs)

- Taximetrics Laboratory

March 29, 1967

Dr. L. R. Hill
National Collection of Type Cultures
Central Public Health Laboratory
Colindale Avenue
London N.W.9
England

Dear Dr. Hill:

Enclosed is a short descriptive statement concerning methodologies which we find useful and practical for biological classification. The statement intentionally does not include much computer program descriptions. We feel that the more important item is the biological description.

If it is desirable to have more descriptive words about the hardware, would you please indicate how this should be done? I trust that this is a satisfactory statement for Taxometrics.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

P.S. Please note that our address will change July 1 from that given on the letterhead to: Taximetrics Laboratory, Department of Biology, University of Colorado, Boulder, Colorado 80302.

Enc.

SIMILARITY-CLUSTERING PROGRAM

This program was developed by the staff of the Taximetrics Laboratory at ~~Colorado State~~ ^AUniversity as an aid for taxonomists in all areas of biology and for workers in other fields as well who have problems in the classification of material, living or otherwise. The following publications explain the methodology in detail.

1. Wirth, M., Estabrook, G., and Rogers, D. J. 1966. A Graph Theory Model for Systematic Biology. *Systematic Zoology*, 15(1): 59-69.
2. Estabrook, G. 1966. A Mathematical Model in Graph Theory for Biological Classification. *Journal of Theoretical Biology*, 12: 297-310.
3. Estabrook, G. and Rogers, D. J. 1966. A General Method of Taxonomic Description for a Computed Similarity Measure. *BioScience*, 16(11): 789-793.

The user selects a collection of objects to be classified and a set of characters which in his professional judgment are critical to the classification of this collection. (For example, if the collection of objects is a group of plant specimens, one desirable character might be leaf insertion. The permitted states of this character might be opposite, alternate, whorled and basal.) The input to the program comprises a description of all objects in the collection. The description of a single object consists of naming that state of each character which applies to this object. The program compares each object, character for character, with every other object and computes a similarity measure for each pair of objects. On the basis of this measure, the program creates ^ahierarchical series of partitions of the collection into clusters of objects. The output is a print of these partitions

together with a measure of the internal affinity of each cluster and its degree of isolation from its nearest neighboring cluster. This printout is not only a possible classification, but provides extremely pertinent summary information which aids the user in making his final taxonomic decisions. The method is particularly useful in situations too complex to be handled by a completely intuitive approach.

The program was written for the IBM 7044 in FORTRAN IV with one subroutine in MAP (7044 assembly language). The program is now being expanded to provide additional useful output.

A program writeup with detailed flow charts and a source program listing are available on request from the authors, as well as reprints of the above publications. Interested parties should contact: Dr. David J. Rogers, Taximetrics Laboratory, Department of Biology, University of Colorado, Boulder, Colorado 80302, U.S.A.

- Taximetrix Laboratory

March 30, 1967

Dr. Donald A. Corlett, Jr.
Scientific Research Department
California Packing Corporation
215 Fremont Street
San Francisco, California 94119

Dear Dr. Corlett:

Thank you for sending the program for identification. If at some future time you decide to document the program, we would be happy to see it.

Under separate cover, I am sending three of our more recent reprints which I hope will be of interest to you.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

*Sent: Bot Science 16(11)
J. Thunb. Bot. 13(1)
Syst. Zool. 15(1)*



California Packing Corporation

AIRMAIL

215 FREMONT STREET • SAN FRANCISCO • CALIFORNIA 94119 • TELEPHONE AREA CODE 415 • SU. 1-7760 • CABLE ADDRESS • CALFRUIT

March 28, 1967

Dr. David J. Rogers
Professor Of Botany
Taximetrics Laboratory
Dept. of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

Enclosed is the program and an example of a list of bacterial isolates which were classified.

As to documentation, Mr. Wolfe and I originally devised this program specifically for the identification and classification of bacterial culture isolates.

I hope it will be of value to you.

Very truly yours,

CALIFORNIA PACKING CORPORATION

Donald A. Corlett, Jr.

Donald A. Corlett, Jr.
Scientific Research Department

DAC:je

Encl.

- Taximetrics Laboratory

February 1, 1967

Dr. D. A. Corlett, Jr.
Department of Food Science and Technology
Oregon State University
Corvallis, Oregon

Dear Dr. Corlett:

Thank you for your reprints "Application of Replica Plating and
Computer Analysis for Rapid Identification of Bacteria in Some Foods,
I and II".

We would be pleased if we might have a copy of the program
developed by Mr. Wolfe which took care of the identification scheme.
Any documentation of the program will be appreciated.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch



California Packing Corporation

215 FREMONT STREET • SAN FRANCISCO • CALIFORNIA 94119 • TELEPHONE AREA CODE 415 • SU. 1-7760 • CABLE ADDRESS • CALFRUIT

February 28, 1967

AIR MAIL

Dr. David J. Rogers
Professor of Botany
Taximetrics Laboratory
Dept. of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

Thank you for your letter of 02/1/67, and I am sorry for the delay. I have written Glen Wolfe a letter and as soon as the program is available, I will send it to you.

This program worked very well for us as a means of searching through a large number of bacterial isolates and placing them in groups, according to a predetermined key.

Very truly yours,

CALIFORNIA PACKING CORPORATION

Donald A. Corlett, Jr.

Donald A. Corlett, Jr., Ph.D.
Scientific Research Department

DAC:mam

MEMORANDUM

March 3, 1967

TO: Dr. Kenneth Medearis, UNIVERSITY COMPUTER CENTER

FROM: David J. Rogers

SUBJECT: For Use in the NSF Computer Proposal

- Project names:
1. Use of computers in classification
 2. Biological information retrieval systems
 3. Information processing systems
 4. Use of IBM equipment in processes of biological classification
 5. A computer-aided classification of Cinchona.
 6. Classification of Cassia
 7. World monograph of marine Isopoda
 8. Classification of Iris missouriensis
 9. Monograph of Manihot esculenta
 10. Classification of sea snakes (Hydrophiidae)

Brief description: The first four projects listed above are attacks on various aspects of the total process of classification. Within these projects are such problems as pattern recognition, similarity-clustering procedures, information processing and retrieval, and several others not yet defined by this group. The work involves the development of rules and logic for the biological problem, the discovery of mathematical models and algorithms, the development of practical and efficient computer programs, and practical testing. The projects listed from five through ten are the applications we make of the programs developed in the first four. These are in conjunction with a variety of biologists around the country, plus our own in-house interests. The projects give feed-back concerning the value of the programs, and corrections to be made in the programs, plus ideas for the development of new methodologies.

Principal Investigators: ROGERS, David J., Professor; FLEMING, H. S., Associate Professor; ESTABROOK, G. F., Instructor; and BRILL, R. C., Programmer.

- Taximetris Laboratory

February 23, 1967

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey 08520

Dear Miss Robichaud:

Your check and letter came on time. I hope you will have a chance to talk to Pierre soon. The work does sound exciting.

Thank you for the information from your College Editorial Department. When I have prepared something on comparative morphology, I will get in touch with them.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

HIGHTSTOWN, NEW JERSEY

BERYL ROBICHAUD

VICE PRESIDENT

609 448-1700

February 16, 1967

Dear Dr. Rogers:

Enclosed is a check for \$150 to cover the programming and computer time spent on the yellow violet data.

I have had no difficulty in analyzing the formation of clusters but now am immersed in the interpretation of the results. When something (or nothing) appears to be of significance, I will let you know. Meanwhile, again thank you very much for the efforts spent on this project. Because I have been away, I have not seen Dr. Dansereau but will talk to him next week about possible future work.

I talked with some people in our College Editorial Department about your project on comparative morphology as a computer assisted project. They suggested that you prepare a writeup detailing the project -- its subject matter, its format in terms of useability on equipment, its cost of implementation and its possible marketability in other universities.

Yours very truly,

Beryl Robichaud

Dr. David J. Rogers
Department of Botany & Plant Pathology
Taximetrics Laboratory
Colorado State University
Fort Collins, Colorado 80521

- Taximetris Laboratory

January 31, 1967

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey

Dear Miss Robichaud:

Just make the check payable to Colorado State University and send it to me. I will have it deposited in the appropriate Colorado State University account so that we can use the money for further computation.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

HIGHTSTOWN, NEW JERSEY

BERYL ROBICHAUD
VICE PRESIDENT

(509) 448-1700

January 27, 1967

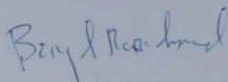
Dear Dr. Rogers:

I have just returned home and unfortunately have not had any time to fully analyze the computer printouts. However, they do make fascinating reading in just the little spare time I have had on airplane travel.

The financing of the project has to come out of my personal pocketbook. For this reason, I prefer to make a check payable to Colorado State University, as you suggested during my first visit. Let me know how this should be forwarded so that it can be credited to the Taxometric Department.

I will be away for the next two weeks and on my return will talk to Dr. Dansereau about the future of the project.

Sincerely,



Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

- Taximetries Laboratory

January 24, 1967

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey 08520

Dear Miss Robichaud:

Enclosed is a statement for computing expenses. I have sent it on University letterhead merely because I do not have any billing forms. To simplify the process, please make checks payable to me.

We enjoyed seeing you and hope that we can continue to work together. I sent off the remaining printout just this morning.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

January 24, 1967

TAXIMETRICS LABORATORY

to

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey 08520

for

Programming and computing expenses. 12 runs of violet data.

Total \$150. Make checks payable to David J. Rogers.

- Taximetries Laboratory

January 11, 1967

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey

Dear Miss Robichaud:

Your data came last week. We have had to do some repunching because (1) either our instructions were not clear or (2) your punch operator added a few extra holes. However, the adjustments are minor and we will do the runs this week. I hope that all the print-out will be back next week and that we can get together to look at it shortly.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

HIGHTSTOWN, NEW JERSEY

BERYL ROBICHAUD

VICE PRESIDENT

(509) 448-1700

December 13, 1966

Dear Dr. Rogers:

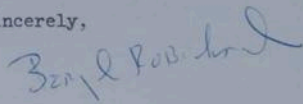
Returning to Hightstown, I find that the potato fields are not nearly as beautiful a backdrop as your mountains in Ft. Collins.

I profited so much from the time which you, Dr. Fleming, and George spent with me. I now feel that I can transmit into input data several tests that could be run profitably with your computer program. I will have to be back out in Colorado in the month of January, so perhaps it could be done at this time.

Meanwhile, thank you all for the time which you spent with me. I know that visiting firemen are not productive for your own research and, therefore, I appreciate the visit.

Happy holidays to all of you in the Taxometric Building.

Sincerely,



Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

McGRAW-HILL, INC.

HIGHTSTOWN, NEW JERSEY 08520

BERYL ROBICHAUD

VICE PRESIDENT

(800) 448-1700

November 28, 1966

Dear Dr. Rogers:

I have rearranged my schedule so as to give you a breathing spell from visiting firemen.

Instead of arriving on Monday, December 5, I will be in Fort Collins by 9:00 A.M. on Tuesday, December 6. Therefore, unless I hear otherwise from you I will make my way to your office early Tuesday. I realize that your time is limited and I will be prepared with data and questions so as to make best use of it.

Sincerely,

Beryl Robichaud

Dr. David J. Rogers
Colorado State University
Department of Botany - Taxonomy Laboratory
Fort Collins, Colorado 80521

Taximetrics Lab.
Nov. 27

Dear Miss Robichaud:

We look forward to your visit December 5. We will be expecting you around noon, if this fits your plans.

Enclosed is a Denver map to get you started. I assume that you will be coming up directly from Stapleton Airport. At any rate, you want to take the Valley Highway, Interstate 25 north. Most of the signs indicate Cheyenne on them. Ft. Collins is 60-65 miles north of Denver.

Your turn-off from 25 to Ft. Collins is Colorado H'way number 14, west. (This is the second of two turn-offs for Ft. Collins.) From the turn-off into the city is about 4 miles. Stay on Colo. 14 without any turns to the 2nd traffic light in town. At this traffic light, is the intersection with US 287. This street is S. College Avenue.

Turn left on S. College, and go two blocks, to the traffic light. This intersection is Laurel Street. The corner of College and Laurel is the northeast corner of the campus of CSU. We are the most northeast building on the campus, and you can see a little, one-story, brick building, (circa 1890), from that intersection. The building is known on campus as the Old Music Building.

Just behind our building is an alley-way, on Laurel St. You can find parking space in the alley.

In case you miss somewhere, call me, and I'll come and get you. Our phone is 491-5201.

Sincerely,

David J. Rogers
Professor of Botany

Encl.-map.

Miss Beryl Robichaud
McGraw-Hill, Inc.
Hightstown, N.J.

Sent 10/21/66 to Miss Rebekah.

"A general method of Saponin
Description for a computed
similarity measure."

- Taxonomy Laboratory

September 7, 1966

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
330 West 42nd Street
New York, N.Y. 10036

Dear Miss Robichaud:

I hope that October can find us in the office sufficiently long so that you may fit us in your schedule. The dates of my absence will be October 14 through 16 (in St. Louis) and October 25 through 30 (at Yale and environs). This leaves, then, in the middle of October, the 10th through the 13th and the 17th through the 21st. I hope this fits your schedule.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

330 WEST 42ND STREET, NEW YORK, N.Y. 10036

BERYL ROBICHAUD
VICE PRESIDENT

NEW YORK (212) 971-2021
HIGHTSTOWN, N.J. (609) 448-1700

September 1, 1966

Dear Dr. Rogers:

I am sorry that I could not arrange to get to Colorado this August. Now, I am off to Portugal for a couple weeks vacation.

It may be possible that I will have to be in Denver some time in mid October. If so, I will let you know to see if it might be convenient if I visit you there. If this does not work out, I would very much like to see you after your affair in New Haven.

Sincerely,

Beryl Robichaud

Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

McGraw-Hill, Inc.

HIGHTSTOWN, NEW JERSEY

BERYL ROBICHAUD

VICE PRESIDENT

800 448-1700

August 4, 1966

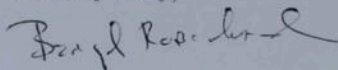
AUG 8 1966

Dear Dr. Rogers:

I hope that you received my message that because of the airline strike it was not possible for me to arrange to get to Denver.

Now, because of other commitments I may not be able to get out until later in the year. If it becomes possible for me to still make a trip this August, I will contact you to see whether my visit would be convenient for you. Meanwhile, if you are planning to come east any time later this year, would you let me know. Thank you.

Yours very truly,



Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

- Taxonomy Laboratory

August 10, 1966

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, New Jersey

Dear Miss Robichaud:

I am sorry that various things have interfered with our getting together on the violets. I will be free the rest of the month with the exception of the 28, 29, and 30, so that if you can make a trip in August these are the only dates I will be unavailable.

I expect to be coming to New Haven and vicinity for a symposium from October 26 through the 28th celebrating the Peabody Museum's 100th anniversary (Pierre will be there). Perhaps we could get together in New York after this symposium.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

330 WEST 42ND STREET, NEW YORK, N.Y. 10036

BERYL ROBICHAUD
VICE PRESIDENT

NEW YORK (212) 971-2021
HIGHTSTOWN, N.J. (609) 448-1700

June 29, 1966

JUL 5 1966

Dear Dr. Rogers:

I have rescheduled my time in Denver in consideration of the rains on the plains in Mexico. I could come and visit you on Thursday, August 4, if this would be convenient.

So, unless I hear from you otherwise, I will appear about 9:30 on August 4.

Yours very truly,

Beryl Robichaud

Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

- Taxonomy Laboratory

July 6, 1966

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, N.J. 08520

Dear Miss Robichaud:

Dr. Rogers has left for Mexico. However, August 4 does not appear to conflict with anything, so we will expect you.

Sincerely,

Henry S. Fleming
Professor of Botany

HSF:ch

June 14, 1966

Miss Beryl Robichaud
Vice President
McGraw-Hill, Inc.
Hightstown, N.J. 08520

Dear Miss Robichaud:

Let me get some dates arranged with you before answering your various questions on the yellow violets.

We have had to shift our time for collecting in Mexico from this month til July, largely because our areas of interest receive rain during that month, and apparently no sooner. Therefore, to make our trip successful, we've had to change our earlier date projections. Does your time in Denver coincide with other activities there to such an extent that you could not delay your visit to, say, July 27-30? We plan to be back here on July 22, and therefore should be available from the 25th onward, but would like to have a few days leeway at the end of our trip, just in case we can't make it back on the stated date. I hope that you can make this rearrangement, because we are quite interested in your project, and want to help out, if we can.

We have discovered that there are many problems involved in readying classificatory data for successful computer runs. Some of these problems are quite subtle, and it is much easier to work them out on a face-to-face basis rather than try to do it by correspondence. So, if you can come on the dates indicated, we will all profit. Bring with you the full set of data, unless you have to have a small truck to carry them. Though you have read our paper in Systematic Zoology, there are other things about our procedures that you should appreciate. Explanation of these can be given during your visit.

Now, to answer specific questions on page 2 of your letter.

1. Yes, the computer has been used for intra-specific studies. This is my own particular problem (with Manihot esculenta), and our programs have shown much merit for such studies.
2. The programs have been designed to accept data of a much larger order than you have indicated has been collected for each violet specimen. The number of pieces of descriptive data for each specimen

is, practically speaking, unlimited. There are obviously theoretical upper bounds, but I do not think any botanist will ever approach the upper boundary.

3. Yes, hybridization will have some effect on the classification, but you can decide what classificatory categories (taxa) you wish to recognize much more readily with the results of a computer analysis than you could by any other means of data correlation. Let us be clear on one point, however. The print-out does not tell you what a particular cluster must be called (i.e., what hierarchical category you wish to assign it). This decision is the professional botanist's and no computer can abrogate it.

4. Our clustering program has a maximum number of objects (specimens) of about 720. Having made this flat statement, however, let me say that there may be ways to overcome this difficulty. First, there are probably many objects with identical descriptors. We have a small program to identify these, and could easily determine whether or not you could eliminate some of them from the study. Secondly, while you may feel that 1500 specimens is a minimum, it might be possible to restrict the number by using your own knowledge of the materials—it is likely that some of the objects lack considerable amounts of information—damaged, badly prepared, failure to include some pertinent data with specimen, etc. These specimens lacking information probably should not be included in the computer study, for lacking information is something the computer cannot do anything with or about.

Computer time is not the really critical element. Suppose with our program you use the computer for 3 hours. Average cost of such computation is \$150.00/hour. \$450 is not absolutely beyond reach for research projects, and this amount can be "found" somewhere. Print-out from 3 hours' run will keep you occupied for months in determining what interpretation you wish to make of it. I might say that we have never run over 230 objects in a single run. This run took about 45 minutes with a much less efficient program than is now running, so would reduce to not more than 20 minutes.

We have not ourselves attempted to do any ecological studies with our program. We are interested to see if there are possibilities, and how they will work, but since we have to depend upon the competence of others for know-how ecologically, we will not attempt to derive and use our own input. Working with those trained in ecology, however, we can suggest useful methods of data handling. There is a group of ecologists working for the Rocky Mountain Forest Experiment Station here, whose interest is in forest ecology, and who are attempting to work with our programs. Their efforts should be of interest to you.

In response to your last paragraph, I think there is definite profit in attempting to use our programs for your work. As earlier stated,

Miss Beryl Robichaud

- 3 -

June 14, 1966

we need to get together to make the project successful. If you could somehow get the rain in the plain in Mexico to come a little earlier, we would have no problems. I hope you can rearrange your schedule.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

McGRAW-HILL, INC.

330 WEST 42ND STREET, NEW YORK, N.Y. 10036

BERYL ROBICHAUD
VICE PRESIDENT

NEW YORK (212) 971-2021
HIGHTSTOWN, N.J. (609) 448-1700

→
Zip 08520

June 8, 1966

JUN 13 1966

Dear Dr. Rogers:

I am the part-time college graduate student who talked to you about the research project I am doing for Dr. Dansereau. I am planning to be in Denver in mid-July and would like to see you on July 12 or July 13 to talk about this project, if you think it worthwhile.

In summary, the research project involves the study of stemmed yellow violets growing in eastern North America involving two aspects -- an analysis of plant specimen characteristics and of colony ecological characteristics.

With respect to the plant characteristics, I have recorded on punched cards data that has been observed and measured on 1501 individual plant specimens. The attachment lists the data recorded on cards. Botanists have tended to use 3 to 8 characteristics to identify species and variety in the stemmed yellow violets. These are

1. Number of stems
2. Number of basal leaves
3. Pubescence - stem and leaf
4. Pubescence - capsule
5. Stipule width-length ratio
6. Number of teeth in stem leaf
7. Shape of stem leaf base
8. Leaf lobing

Dr. Dansereau is suggesting that there are two species of yellow violets in the collection, each of which has two varieties. However, because the stemmed yellow violets appear to hybridize freely, it is unusual to find "pure" plants in each of the four

groups. Rather, only by using the average characteristics of all plants in a given colony, can one identify tendencies of colonies to be more like one species than the other.

I have used the punched card specimen data to construct several different indices (using the principle of the Anderson Index of hybridization) for analysis of plant characteristics by colony. Combinations of the eight characteristics mentioned above have been used for the indices.

My question now is whether or not it is appropriate to use computer programs such as those developed by you (and I have read your recent article in Systematic Zoology) to analyze the plant specimen data that I have recorded. Specifically,

- Has the computer been used for any comparable study involving species and variety classifications?
- Is there merit in analyzing by computer all of the specimen data listed on the attachment inasmuch as botanists generally to date have concluded that only certain characteristics are important?
- If the plants have hybridized freely, will this not affect the classification?
- Is it impractical to analyze as many as 1501 specimens? I had gathered from your previous writings and the Sokol book that because of the amount of computer time, it is practical to analyze only a limited number of specimens, yet the value of this study rests with analysis of a large number of specimens.

I am now in the process of recording on punched cards ecological data about each of the colonies in which the collection of stemmed yellow violets were made. Among the data here is a listing of other species (and I have noted these by number) that are found growing with the stemmed yellow violet. With respect to this data,

I am wondering if you or your colleagues have done any work in using a computer to compare colonies on the basis of ecological characteristics such as associated species.

If you feel that there would be no profit in my exploring further the possible use of a computer to analyze the yellow violet data, please do not hesitate to let me know. If, on the other hand, you think that by use of the computer we might throw more light on the taxon of the stemmed yellow violet, I would like to pursue with you further the study. I could be in Fort Collins on July 12 or 13th to talk personally with you about the study. Would you let me know if you think it worthwhile?

Yours very truly,

Beryl Roachland
Beryl Roachland

Attachment

Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

Characteristics of Plant Specimens of taxa of
Stemmed Yellow Violets of Eastern North America

Data is recorded on 80-column punched cards as follows:

<u>Characteristic</u>	<u>Card Column</u>	<u>Item of Data</u>
Reference	1 - 2	Colony Number
Reference	3 - 4	Plant Specimen Number
1	5	Number of Stems Cards now contain values of 1 to 8 but should be reduced to three values - 1, 2 and 3
2	6	Stem pubescence Cards now contain codes for four values but should be reduced to three values - 1, 2 and 3
3	7 - 9	Length of Stem Data recorded in MM and range is from 043 - 372 MM
Ignore	10 - 11	Position of bract (Characteristic is of no value)
4	12	Number of Stem Leaves Cards now contain codes for 9 values but should be reduced to 4 values
5	13	Pubescence Stem leaf (Ventral Size) Cards contain four values - 1, 2, 3, 4
Ignore	14	Pubescence Stem leaf (Dorsal Side) No value
6	15 - 16	Length of Stem Leaf (MM) - Actual Measurement
7	17 - 18	Width of Stem Leaf (MM) - Actual Measurement
8	19	Shape of base of Stem Leaf expressed in three values - 1, 2, and 3
9	20 - 21	Number of Marginal Teeth in Stem Leaf
Ignore	22 and 23	Stipule Dorsal & Ventral Pubescence - No value

<u>Characteristic</u>	<u>Card Column</u>	<u>Item of Data</u>
10 ?	24 - 25	Length of Stipule (MM) Actual Measurement
11 ?	26	Width of Stipule (MM) Actual Measurement
Ignore	27	Stipule Margin Pubescence
12	28	Basal Leaves - Number
		Cards are punched with values - 0, 1, 2, 3, 4, and 5. Should be reduced to 0, 1, 2, and 3.
Ignore	29 - 37	Basal Leaf data common only to some plants - No Value
13	38	Capsule Pubescence
		Cards now punched with values 1, 2, and 3
Ignore	39 - 40	First stem internode ratio - Incomplete $\frac{L}{W}$
14	41	Stem Leaf lobing - Cards punched with values 0 and 1
15	42 - 44	Stem Leaf shape Ratio (Width/Length) 2 decimals
16	45 - 47	Stipule Shape Ratio (Width/Length) 2 decimals
Ignore	48 - 50	Basal Leaf Size Ratio (Width/Length)
Index Conversions and calculations punched in columns 51 - 80		

Analysis of Taxon of Stemmed Yellow Violets
Characteristics and Number of Possible Values

<u>Characteristic</u>	<u>Number of Values</u>	<u>Item</u>
1	3 (or 4)	No. of Stems
2	3	Stem pubescence
3	? 9	Length of Stem
4	4	No. of Stem leaves
5	4	Stem leaf pubescence
6	? 9	Length of Stem leaf
7	? 9	Width of Stem leaf
8	3	Shape of base of stem leaf
9	? 9	No. of teeth in stem leaf
10	? 9	Length of Stipule
11	? 9	Width of Stipule
12	4	Number of basal leaves
13	3	Capsule pubescence
14	2	Lobing of stem leaf
15	? 9	Stem leaf shape ratio
16	? 9	Stipule shape ratio

- Taximetrics Laboratory

February 2, 1967

Dr. Azriel Rosenfeld
Computer Science Center
University of Maryland
College Park, Maryland

Dear Dr. Rosenfeld:

Your recent paper "Sequential Operations in Digital Picture Processing" which appeared in J. A.C.M., Oct. 1966, 13:4 interested me very much, for reasons essentially unrelated to digital picture processing. I am especially interested in the program mentioned near the bottom of page 488, which discovers proximity graphs, computes the order of connecturty for each node and actually draws the components. I would be most interested in any descriptions or other documentations (as well as any comments you would care to make) related to this program.

Thank you for your cooperation.

Very truly yours,

George F. Estabrook

GFE:ch



DEPARTMENT OF PREVENTIVE
MEDICINE AND PUBLIC HEALTH
BIostatistical UNIT AND
MEDICAL RESEARCH
COMPUTER CENTER

THE UNIVERSITY OF OKLAHOMA
MEDICAL CENTER

800 NORTHEAST THIRTEENTH STREET
OKLAHOMA CITY, OKLAHOMA, 73104

February 1, 1967

Dr. G. F. Estabrook
Department of Botany
and Plant Pathology
Colorado State University
Fort Collins, Colorado

Dear Dr. Estabrook:

In your very interesting paper entitled "A Mathematical Model in Graph Theory for Biological Classification", you indicate that "standard FORTRAN listings are available upon request" for a program embodying your approach. I would very much appreciate receiving such a listing if possible.

Please remember me to my old friend Frank Graybill if you come into contact with him.

Thank you very much.

Sincerely,

Edward N. Brandt, Jr., M.D., Ph.D.
Director, Computer Facility
and Biostatistical Unit

ENB/rt

- Taximetrics Laboratory

February 8, 1967

Dr. Edward N. Brandt, Jr.
Director, Computer Facility & Biostatistical Unit
University of Oklahoma Medical Center
800 Northeast Thirteenth Street
Oklahoma City, Oklahoma 73104

Dear Dr. Brandt:

Please find enclosed the listing of the program which you requested. The chart and appendices referred to in the associated comments are, unfortunately, in short supply presently. However, the program has been submitted to SHARE Library and complete documentation will be available from SHARE soon. I am sorry I cannot be of more immediate assistance at present. In an effort to meet the spirit of your request, I have enclosed two documents describing the input to the program--aspects of the total problem not dealt with specifically in the Journal of Theoretical Biology publication. I hope you find them helpful as well.

Very truly yours,

George F. Estabrook

GFE:ch

February 14, 1967

Dr. W. Hardy Eshbaugh
Department of Botany
Southern Illinois University
Carbondale, Illinois 62901

Dear Dr. Eshbaugh:

Thank you for sending a copy of your dissertation abstract. It is difficult to know precisely what comments can be made concerning your dissertation.

Someshow or other, however, I feel that to say that "taximetrics" will not work (or does not agree with the cytogenetic method) is not appropriate. Also to say that the "cytogenetics" method differs from the "taximetric" method does not tell me anything unless I know precisely what is meant by both.

I can sympathize with you when it comes to differentiation of cultivated species from wild ones or vice versa. We have this difficulty in Manihot. However, I would not come to the conclusion that there is some valid difference between the species because of the manipulation of the data according to one or another scheme.

We feel very strongly that one has to do a considerable amount of work with one's input information, "characters", in order to reflect the desires of the taxonomist. After all you determine what your output should look like, not the computer.

Now, having said these things, I would like to know if you would be interested to try out the methodology we have recently developed. If so, we should get together to be certain each understands the other's problem, to know which taximetric method is appropriate. There is no one taximetric method which will satisfy all considerations. Let us see if ours has any merit for your problem.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

Southern Illinois
University

CARBONDALE, ILLINOIS 62901

February 10, 1967

Department of Botany

Dr. David J. Rogers
Taximetrics Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

I am enclosing a copy of my Abstract with reference to my thesis. The numerical taximetric treatment used in my study was accomplished prior to the publication of Dr. Sokal's book. Before I could consider publishing this material, I would have to rework my data using some of the more recent methods that have appeared in the literature.

My study was intended to compare quantitative morphological data with known genetic data. My preliminary study indicates that cultivated species and their known wild progenitors should not be studied in the same taximetric system. In fact, Sokal's numeric system ^{formed} ~~divided~~ the lowest possible correlation for a wild species and its known derived cultivated species.

Should you wish any further information on this subject, I would be happy to correspond with you. I also would welcome your comments on this matter.

Sincerely yours,

W. Hardy Eshbaugh
W. Hardy Eshbaugh
Assistant Professor

WHE:bh

encl.

A NUMERICAL TAXONOMIC
AND CYTOGENETIC STUDY OF
CERTAIN SPECIES OF THE GENUS CAPSICUM

(Order No. 64-12,019)

William Hardy Eschbaugh III, Ph.D.
Indiana University, 1964

The relationship of several taxa of the genus Capsicum (Solanaceae) has been studied using both cytogenetic and taximetric methods.

Taximetric relationships are established by using product-moment coefficients of correlation obtained by Q-technique of factor analysis. Group relationships among the taxa are found by the variable-group method and the ramifying-linkage method of cluster analysis. Cytogenetic relationships are based on data from hybridization studies, pollen stainability, karyotype analysis, meiotic chromosome behavior, and F_1 and F_2 morphology.

Two species of Capsicum, the cultivated C. pendulum and the wild C. microcarpum were studied in detail. Qualitative characters show these two taxa to be relatively inseparable while quantitatively they are quite distinct.

Taximetrically the association of these two taxa is very low. However, they form fertile hybrids which generally exhibit normal chromosome pairing, and they also have the same karyotype. It is thought that the cytogenetic information is more revealing of the true relationship of these two taxa. Therefore, it is proposed that these two taxa be combined into a single species and given varietal status. The wild taxon C. microcarpum then becomes C. baccatum var. baccatum and the cultigen is known as C. baccatum var. pendulum (Willd.) Eschbaugh. The role of cultivation in the evolution of C. baccatum var. pendulum the taximetric treatment of cultigens and the nomenclature of these two taxa are discussed.

An additional study of the taximetric and cytogenetic relationships of fourteen taxa, including the two above named varieties, was made. The variable-group method is quite conservative and only reveals two groups while the ramifying-linkage method is more liberal and depicts nine distinct groups at the same level of analysis. The relative merits of the two techniques of group analysis are discussed.

Cytogenetic evidence suggests that there are two distinct groups among these fourteen taxa. The first group consists of six cross fertile taxa, C. cardenasii, C. eximium, SA 321 - SA 357, C. pubescens, SA 270, and SA 281, that form highly fertile F_1 hybrids. The second group comprises five cross fertile taxa, C. baccatum, C. praetermissum, C. frutescens, C. chacoense, and C. chinense, that produce sterile progeny from intercrosses.

A comparison of the group structure obtained from taximetrics and cytogenetics is made. The ramifying-linkage method reveals the same two groups found by cytogenetic methods, as well as seven additional groups. The relationship of all these groups could not be tested by hybridization studies because of a lack of flowering material. The variable-group method depicts only one of the groups revealed by the cytogenetic evidence. Furthermore, it would seem that numerical taxonomy as now developed is inadequate for the treatment of both cultivated and wild species in a single system.

Several taxa used in the present study represent undescribed species and will be treated formally elsewhere.

Microfilm \$2.75; Xerography \$6.00. 124 pages.

- Taximetries Laboratory

January 26, 1967

Dr. Phyllis V. Parkins
Director, Biological Abstracts
3815 Walnut Street
Philadelphia, Penn. 19104

Dear Phyllis:

We will soon submit to Dr. Adkinson's group in NSF a proposal for the development of an IR system for biologists. Our take-off point is the fact that many folks have worked on IR systems for literature but very few have made efforts towards IR systems for raw data. As you are aware, the two ideas do not necessarily duplicate each other.

We are hoping, however, to be able to ask you (or some of your assistants) to give us some time and advice on the development of such a system. We feel that there is no more knowledgeable source than yours about biological information retrieval and that we could profit muchly by talking to you and your programmers. I won't go into any detail about our hoped for system because it would require considerable amount of time to explain our ideas. I am sending under separate cover copies of our grant request to NSF so that you can get an idea of what we are up to.

If the grant proposal is approved, we hope to get started next summer. If we do get started then, we would hope to be able to talk to you folks some time shortly after the beginning period of the grant. I hope that we can count on your aid and advice.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

BioSciences Information Service
OF BIOLOGICAL ABSTRACTS

MEMO

To: All Section Editors

From: Phyllis V. Parkins, Director *PH*

Subject: A plan to strengthen our lines of communication

During the past two years or so circumstances have made it increasingly difficult to communicate with many of you as freely or as frequently as I would have liked. But I do hope you understand how much all of us at BA, and members of our editorial staff in particular, appreciate the efforts you are putting forth on behalf of BA. We especially appreciate your willingness to try out and adapt to new procedures from time to time, as we have acquired modern electronic equipment, and have been able to take new approaches to deal with the even greater quantity of literature that keeps coming through our doors.

Because these editorials tend to reflect activities and thinking at BA, we believe that you as a contributing member of the BA "family" would be interested in seeing many of them. We are placing you on our mailing list, therefore, to receive reprints as they become available. And as a hint of what is to come, we are enclosing copies of several of those that have appeared this year. We hope you will welcome this opportunity to keep in closer touch with affairs at Biological Abstracts.

Have you perchance come across two articles about BA that have been published during the current year? The first paper, describing our overall operations, appeared in Science, May 13; the second, describing an experimental search service, was printed in the September number of BioScience. We still have a few reprints on hand, should you be interested in having them.

And now--to each of you my very cordial good wishes for a happy Christmas and an especially rewarding New Year.

*P.S. Just to let you know - we're beginning to do
unto others as we have been doing unto you.*
P

December 16, 1966



BioSciences Information Service
OF BIOLOGICAL ABSTRACTS

3815 WALNUT STREET PHILADELPHIA, PA. 19104 U.S.A.

Telephone: (215) 386-0414 • Cable BIOSIS Philadelphia • TWX 215-569-8821

Executive and Editorial Offices • PHILADELPHIA
Latin American Editorial Office • MEXICO CITY

August 24, 1966

BOARD OF TRUSTEES

ROBERT J. DODGE
1950-1954

PAUL A. MILES, JR.
1954-1958

ROBERT A. STONE
1958-1962

ALFRED E. THOMSON
1962-1966

PAUL J. FISHBEIN
1966-1970

WILLIAM C. STONE
1970-1974

WILLIAM C. STONE
1974-1978

PHYLIS V. FARKINS
1978-1982

ROBERT A. STONE
1982-1986

Dear Gene:

Last week during the AIBS meetings someone brought up a subject near and dear to my interests--the work of the NAS/NRC Committee on the Use of Electronic Computers in the Life Sciences. It reminded me that I have long intended to drop you a note, suggesting that the Committee in question appears to lack representation from the taxonomic or ecologic areas of biology, both of which would benefit enormously, I believe, from the proper application of computers to research.

Should it be possible, at some future date, to add to this Committee's membership, you might want to consider David J. Rogers, Professor of Botany at Colorado State University. For several years Dr. Rogers has been actively interested and has gained considerable experience in the computer-processing of taxonomic data. In any case, I would imagine that Roger's approach to the use of computers in research would be of interest to the Committee, if they are not already aware of his work.

It seems ages since we've met at one or another official gathering. I do hope all is going extremely well for you.

With kindest regards,

Sincerely,

Phyllis V. Farkins
Phyllis V. Farkins
Director

Dr. C. E. Sunderlin
Special Assistant to the President
National Academy of Sciences
2101 Constitution Avenue
Washington, D. C. 20418

FVP:nmp

bcc: Dr. David J. Rogers ✓

- Taxonomy Laboratory

June 13, 1966

Dr. Phyllis V. Parkins
Director, Biological Abstracts
3815 Walnut Street
Philadelphia, Pennsylvania 19104

Dear Phyllis:

Bill Steere recommended recently that I make an effort to get onto the National Academy of Science's - National Research Council Committee on the Use of Electronic Computers in the Life Sciences. After looking over a list of the members of that committee, I highly agree that it should have representation from one of the members of biology for the more or less classical schools. As presently constituted, there are nothing but physicists and chemists on that committee, and apparently no member who represents taxonomy or ecology, or any of the "classical" parts of biology.

I wonder if there is any possibility of your putting a word in the right place to help out in getting me onto that committee? If there is any embarrassment about this to you, please do not hesitate to let me know, but I will certainly appreciate it if there is anything that you can do to put a word in the right place. Thank you very much.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

DRAKE UNIVERSITY
DES MOINES, IOWA 50311

January 13, 1967

Professor David J. Rogers
Professor of Botany
Colorado State University
Fort Collins, Colorado 80521

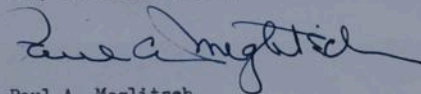
Dear Professor Rogers:

I am at the moment collecting papers and ideas preparatory to trying some numerical taxonomy. It sounds very much as though the technic used in the science article is not one that I should start with. As I don't expect to be a historian in this field, I am sure I am better off with the papers you sent.

I suspect that a good deal of numerical work in taxonomy is going on in Colorado State since the technics and principles that are useful in one group of organisms are so often useful in other very remotely related creatures.

I would appreciate having copies of papers published from the taxonomists laboratory. I am send one or two papers of my own which may have some general interest to you. I would be glad to send copies of systematic papers that may come out of the Drake Program.

Very sincerely yours,



Paul A. Meglitsch
Professor of Zoology
Biology Department

PAM:djn

- Taximetrics Laboratory

January 4, 1967

Dr. Paul A. Meglitsch
Biology Department
Drake University
Des Moines, Iowa 50311

Dear Dr. Meglitsch:

Your letter requesting reprints of our paper published in Science has been forwarded to me at the above address. I will be glad to supply this paper if after the following explanation you would still like to have it.

As far as we are concerned the paper is now of historical significance only and does not represent our latest recommended procedure for clustering. First of all, it is an extremely complex and cumbersome procedure that does not lend itself well to computer methodologies. While the procedure would run on IBM 704 (and as I understand was later reprogrammed for a 7090 in England), it still takes a monumental amount of time and the results are not particularly satisfying.

We do recommend the program described in the enclosed paper from Systematic Zoology. We have had a number of test cases on both plants and animals and find it very useful. I recommend strongly that you read the companion piece that I am also sending. This paper, published in BioScience this past year (1966), is necessary to understand the procedures of preparing data for the clustering method described in the reprint from Systematic Zoology.

If after this you care to still have a copy of the paper from Science, I will have a Xerox copy made for you.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

DRAKE UNIVERSITY
 DES MOINES, IOWA 50311

DEC 19 1966

December 12, 1966

Mr. D. Rogers
New York Botanical Garden
New York 58, New York

Dear Mr. Rogers:

I should like to have a copy of your paper entitled "A Computer Program for Classifying Plants" from Science. I am engaged in applying computer methods to some taxonomical problems involving Myxosporidia. If you have other publications that you think might be of interest to me, I would be very happy to have copies of them also.

Would you wish to engage in a regular reprint exchange of papers dealing with computer taxonomy.

Very sincerely,

Paul A. Meglitsch

Paul A. Meglitsch
Professor of Zoology
Biology Department

PAM:djn

January 17, 1967

Miss Amy Jean Gilmartin
Department of Botany
University of Hawaii
Honolulu, Hawaii 96822

Dear Miss Gilmartin:

I am unable (I'm sorry to say) to uncover our previous correspondence and so cannot recall the questions in your earlier letter. I think from the statement that you made in your recent letter, however, that I can get an idea of what some of these questions were. If I do not answer all of them, I trust that you will forgive me and ask the questions again.

Included herewith is a reprint of a paper published in Systematic Zoology describing the new graph theory clustering method. The graph theory clustering method is so superior to the earlier information theoretic model that we have completely discarded the latter. In a number of tests on a variety of organisms, both plant and animal and at hierarchical levels from family down to intraspecific variations, this model has proved its worth. Several papers with applications are now in press and should be available soon. The 1966 BioScience paper by Estabrook and Rogers describes the method of input to this new clustering method.

We have not salvaged any portion of the earlier information theoretic clustering model. While there were probably some excellent ideas in it, the difficulty in conceptualizing programming of the model itself were too complex. It is too bad if you are 2/3 of the way through the program of that model and still expect to get any good from it. It simply is not worth the effort.

If you want, we will give you a hand with some of your work, and run some tests for you. With the latest BioScience paper in hand you should have little difficulty in understanding how to set up characters for a run. Did you ever tell me the nature of your project? What size of input do you have? How many objects, how many attributes and characters?

Please pay close attention to our statement that no computer (or model) can dictate to you what information you must use to make a

Miss Amy Jean Gilmartin

- 2 -

January 17, 1967

classification. You cannot abrogate your responsibilities, you must make the choice. The old statement "garbage in, garbage out" cannot be too much emphasized. If you do not have 60 or 50 or some other dictated number of characters, please do not manufacture them to make that number. Use only those pieces of information that you as a taxonomist know to be valid and interesting to you. No statistical technique will tell you what is good or bad information. Please listen to these words!

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

Dr. David Rogers
Department of Biology
Colorado State
Fort Collins
Colorado

Dear Dr. Rogers,

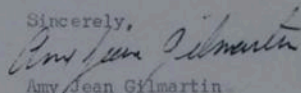
Thankyou very much for the reprint of the recent paper from BioSci.
Am I correct in assuming that since none of the questions in my letter
were answered that this indicates your's and Flemming's complete dis-
illusionment with the method put forth on distance function and your
cluster analysis?

Since I have programmed for this method with some slight modifications
and am about 2/3 of the way through the thesis naturally I will probably
continue using this method at the present. Are there any portions that
you have since discovered which bear close scrutiny?

I really would sincerely appreciate receiving your comments even if you
can not afford to take the time to answer all of my questions.

Again, thanking you for the reprint, I will be hoping to hear from you,

Sincerely,



Amy Jean Gilmartin
Department of Botany
University of Hawaii
Honolulu, Hawaii 96822

P.S. Ted Crovello, presently at KU has offered to run some of my data
through their system which uses correlation rather than distance. I
plan to take him up on this. You probably noticed the mention of the
paper by Proccina and Marks in BioSci. of their work in using both
distance and Pearson's product moment correlation?