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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

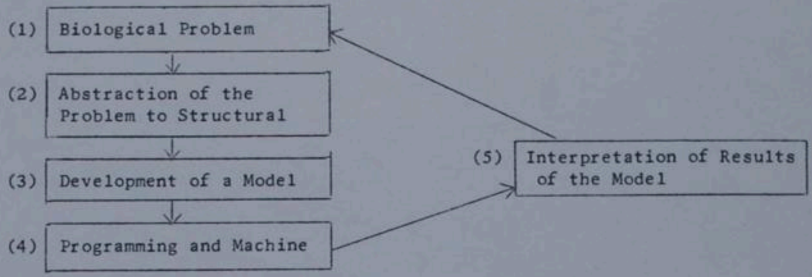
Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do 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.

B 170
April 6, 1966
Mr. George F. Estabrook

## I. Why Learn About Math.



Areas (2) and (5) are the most important reasons why biologists should know about structural mathematics because here is where dialogue must exist between the biologist whose problem is in question and the mathematicians and programmers whose attempt it is to address this problem with the aid of a computing machine. Failure to get from (1) to (2) and from (5) back to (1) is the reason many attempts by biologists to use computers have proved fruitless and frustrating.

Example of (1) → (2) difficulty is Camin-Sokal Evolution

(5) → (1) is some of the statistical approaches which prove difficult to interpret.

The experience and training required to adequately perform operations (3) and (4) properly does not in general belong to a biologist who has spent his time, effort, and energy preparing himself in his own subject. Therefore, we will try to learn something *about* mathematics rather than learn mathematics per se. We are trying to develop an attitude and appreciation so that mathematical collaboration can occur in the appropriate manner from the appropriate source.

## II. The Number System:

### A. Integers

1. Peano - mathematician who invented axioms for positive integers.
    - i. There is some number called a unit.
    - ii. For each number x there is a unique successor x', i.e., if x = 5 then x' = 6.

2. The group structure of integers.

   i. The operation "+" is:

      a. Binary, i.e., operates on pairs of integers.

      b. Closed, i.e., the result of the operation "+" will also be an integer.

   ii. The existence of a"zero."

      a. The "zero" is that number "0" which satisfies $x + 0 = x$ for all integer values of $x$.

      b. Zero is unique. Assume there are two integers $0$, $0^*$ with property "a". Then the "battle of the zeros" gives

$$0^* = \boxed{0 + 0^*} = 0$$

   iii. Solvability.

      a. $x + ? = 0$ must always have an integer solution, called additive inverse of $x$.

      b. This means that minus integers, e.g., $-17$, are also integers.

Note: Any collection of objects* with properties i, ii, iii is called a group.

3. Multiplication on integers.

   i. Multiplication is also closed - binary.

   ii. Its "zero" is actually the unit 1.

   iii. It is not solvable. $x$ times?? = 1 has for solution $? = x^{-1}$ called the multiplicative inverse of $x$. $x^{-1}$ is not always an integer. If $x = 0$, $x^{-1}$ does not exist.

B. Collections of numbers larger than the integers.

1. Rationals.

   i. Rational number is defined to be any number of the form $a$ times $b^{-1}$ where $a$ and $b$ are integers and $b \neq 0$.

   ii. Rational numbers are ordinary fractions.

---

*Object is a general term which also includes concepts, ideas, events, etc.

C. There are other even larger collections of numbers(whose definitions are difficult).

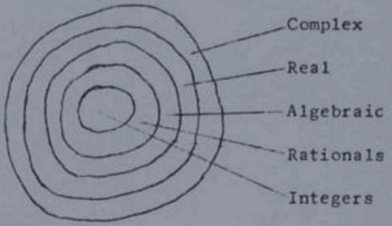
1. For example.

   i. Algebraic numbers

   ii. Real numbers

   iii. Complex numbers

2. The nesting relation for these collections of numbers is shown:



3. The rational numbers is the smallest "ordinary" collection of numbers where addition and multiplication (together with inverse operations) is closed. The <u>vast</u> majority of our problems and thinking in taxonomy will not call for larger collections of numbers than the rationals.

D. Integer arithmetic and representation.

1. Base 10 system.

   i. Digits of a number can be thought of as the coefficients of the expression

$$\sum_{i=0}^{N} C_i \, 10^i \qquad \text{where } N + 1 \text{ is the}$$

number of digits. For example the number 3 4 6 can be written $3 \times 10^2 + 4 \times 10^1 + 6 \times 10^0$.*

---

* A number raised to the $0^{th}$ power is always 1.

2. In the base b system a number $C_2 \quad C_1 \quad C_0$ is a shorthand way of writing $C_2 \times b^2 + C_1 \times b^1 + C_0 \times b^0$.

3. Conversion from base a to base b can be done as follows:

   i. Example 1, base 10 → base 2. Convert 197 to the equivalent base 2 number.

   Step 1. Divide 197 by 2 to get 98 with a remainder of 1.

   Step 2. Divide 98 by 2 to get 47 with a remainder of 0.

   Continue until there is nothing left to divide into as shown:

   197 / 2 = 98 R = 1
   98 / 2 = 49 R = 0
   49 / 2 = 24 R = 1
   24 / 2 = 12 R = 0
   12 / 2 = 6 R = 0
   6 / 2 = 3 R = 0
   3 / 2 = 1 R = 1
   1 / 2 = 0 R = 1

   The remainders arranged in reverse order, i.e., 11000101 is the desired binary number.

   To convert back simply use the definition in D-2, i.e.,

$$1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$

$$= 128 + 64 + 0 + 0 + 0 + 4 + 0 + 1$$

$$= 197$$

   ii. Example 2, convert from base 5 to base 8. Convert 430201 to the equivalent base 8 number. We will use the same algorithm (only now we must do arithmetic in base 5 not 10).

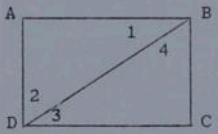
   430201 / 13 = 24203 R = 2
   24203 / 13 = 1400 R = 3
   1400 / 13 = 103 R = 1
   103 / 13 = 3 R = 4
   3 / 13 = 0 R = 3

   The desired base 8 number is 34132.

E. Integers as names.

1. Frequently, integers are used for naming things that don't have other natural names. We will be doing this often throughout the course. It is important not to confuse the numerical properties of a number with its mnemonic properties.

   i. Fallacies: geometry



if <1 = < 3 and
<2 = < 4 prove
<ADC = <ABC.

Proof. < ADC =< 2 +< 3 = <5
< ABC =< 1 +< 4 = <5

<ADC = <ABC.

Don't let this happen to you! .

## III. Set Theory.

### A. Definition.

1. A set is any collection of objects.*

2. The objects in a set are called <u>members</u> of the set.

3. A set with no members is called empty.

4. A set can be specified in one of two ways:

   i. Providing a rule which tells whether or not something is a member, e.g., {** x / x is an integer greater than 3 and less than 8 }.

   ii. Enumerating the membership, e.g., the set described above can also be written { 4, 5, 6, 7, }.

5. The symbol "ε" means "is a member of." $a \varepsilon A$ means a is a member of the set A.

6. The symbol $\subset$ read "is contained by" is defined as follows: $\underline{A}\subset\underline{B}$ means whenever $x \varepsilon \underline{A}$ then x is also in $\underline{B}$. $\underline{B}\supset\underline{A}$ is read B contains A and means whenever $x \varepsilon A$ then x is also in B. If $\underline{A}\subset\underline{B}$ then we say that A is a subset of B. A set is always a subset of itself. The empty set is a subset of every set.

### B. Operations.

1. $\cap$ read "intersection."

   i. Defined as follows: $x \varepsilon A\cap B$ means $x \varepsilon A$ <u>and</u> $x \varepsilon B$. $A\cap B$ is then the set of members common to both A and B.

   ii. This is a binary operation.

   iii. Whenever we consider the set of all the subsets of some large set, (universal set) for this set of subsets $\cap$ is also closed.

---

* Same note on objects, page 2.

** { Will be used to denote a set.

iv. Question: is the set of all subsets of some universal set together with $\cap$ a group?

  a. The zero for this operation is the universal set, call it U.

  b. $A\cap ? = U$ does not always have a solution for ?.

  c. Answer: NO

2. $\cup$ read "union."

  i. Defined as follows: $x \varepsilon A\cup B$ means $x \varepsilon A$ or $x \varepsilon B$ or both.

  ii. Similar to intersection $\cup$ is a closed binary operation when we are working with the subsets of some universal set.

  iii. Is all the subsets of some universal set together with $\cup$, a group?

3. Consider the subsets of some universal set [which is generally the case, and which will be assumed here often unless otherwise noted].

  i. The symbol ~ when written above the name of a set is read "complement of ________ ← name of the set goes here."
  $x \varepsilon \tilde{A}$ means $x \varepsilon U$ but x is <u>not</u> in A. [$\notin$ will mean "not a member of].

  ii. The symbol — read "but not" is defined as follows:

$$x \varepsilon A - B \text{ means } x \varepsilon A \text{ and } x \notin B.$$

C. Boolian algebra is similar to arithmetic only sets are used instead of numbers.

  1. Examples of some operations in Boolian algebra are given:

  i. $\widetilde{(A\cup B)} = \tilde{A}\cap\tilde{B}$

  ii. $(A\cup B)\cap C = (A\cap C)\cup(B\cap C)$

D. Relations.

  1. A relation can be thought of as a rule which tells whether or not some ordered pair of members of some set are "related" according to some criterion.

i. For example, consider the set of living people. Let B stand for "is brother of." We write a B b to indicate that person a is the brother of person b.

ii. Another example: Consider integers. Let D mean "is evenly divisible by". Then we write 12 D 4 but not 13 D 3, further we do not write 4 D 12.

2. Kinds of relations.

   i. A relation R is symetric if a R b indicates that b R a as well.

      a. For example: B in D1i is not symetric, for if I am someone's brother, then that person my be my sister.

      b. A (aAb means $a+b = 3$) is a symetric relation on integers for $a + b = 3$ indicates that $b + a = 3$ as well.

   ii. A relation is reflexive if a R a is always true.

      a. < read "is less than" on integers is not reflexive for $a < a$ is not always true.

      b. $\leq$ read "less than or equal to" is reflexive for $a \leq a$ is always true.

   iii. A relation is transitive if a R b and b R c indicates that a R c as well.

      a. T read "is taller than" on people is transitive.

      b. B read "has beaten" on a league of football teams is not transitive.

3. Equivalence.

   i. Definition. A relation is called an equivalence relation if it is symetric, transitive and reflexive.

   ii. For example:

      a. "=" read equals on integers is an equivalence relation.

b. M read has "the same mother" is an equivalence relation.

c. W read "within a mile of" on houses is not an equivalence relation.

iii. An equivalence relation, then, indicates whether two things are "equivalent" with respect to some criterion. An equivalence relation on some set of objects serves to divide the set into exclusive, exhaustive subsets called "equivalence classes under the relation."

iv. Any scheme for dividing a set into exclusive, exhaustive subsets is called a partition; the subsets into which the origional set was divided according to the scheme are called the classes of the partition. Any partition determines an equivalence relation; namely, I, read "is in the same class as". Satisfy yourself that I is an equivalence relation.

## IV. Logic.

A. Propositional calculus.

1. Connectives.

i. Definitions

a. $\sim$ read "not"

b. $\wedge$ read "and"

c. $\vee$ read "or"

d. $\rightarrow$ read "then" or "implies"

e. $\leftrightarrow$ read "is" or "if and only if".

ii. A proposition, symbolized usually with a small letter from the middle of the alphabet, i.e., p q r s t, etc., is a statement which is truth functional with respect for some set of situations; that is, if a situation from this set is designated, the truth (or falsity) of the statement in question can be unequivocally determined.

a. If $p$ is a proposition, then $\sim p$ is true whenever $p$ is false and $\sim p$ is false whenever $p$ is true.

b. $p \wedge q$ is true whenever both $p$ and $q$ are true.

c. $p \vee q$ is true whenever either $p$ or $q$ is true.

d. $p \rightarrow q$ is false only when $p$ is true and $q$ is false.

e. $p \leftrightarrow q$ is true whenever $p$ and $q$ are <u>both</u> true or both false.

iii. A statement which is always true is called a tautology. An example would be: $(p \rightarrow q) \leftrightarrow \sim(p \wedge \sim q)$.

2. Truth tables.

i. Since the definitions of the connectives above are all in terms of when something is true or false, it is possible to build tables that show when certain compound statements are true.

a. If one proposition is in question, it can be either true or false. Its table looks like:

| p |
|---|
| T |
| F |

b. The table for ~p looks like:

| ~ | p |
|---|---|
| F | T |
| T | F |

Under the connective is placed the truth value for the new proposition (consisting of the old proposition together with the connective).

c. Whenever two propositions have been combined with a connective to form a single proposition, there are 4 possible combinations of basic truth. These are shown below.

| p | ∧ | q | = | R |
|---|---|---|---|---|
| T |  | F | = | F |
| T |  | T | = | T |
| F |  | F | = | F |
| F |  | T | = | F |

By definition Aib2 the new proposition R is true as shown. By convention the truth values of the compound proposition are entered under the connective.

d. Further examples.

i.

| p | v | q |
|---|---|---|
| T | T | F |
| T | T | T |
| F | F | F |
| F | T | T |

ii. p → ~ q

| p | → | ~ | q |
|---|---|---|---|
| T | F | F | T |
| T | T | T | F |
| F | T | F | T |
| F | T | T | F |

iii. Complete the truth table for the compound proposition shown.

p ↔ q

B. The analogue of propositional calculus in Boolian algebra.

1. The universal set for the Boolian algebra to be discussed here is the set of situations IVA1ii over which the propositions to be considered are truth functional.

   i. To each proposition, p, there corresponds some subset of the universal set. If A⊂U corresponds to p then a ε A ↔ p is true for situation a.

   ii. If A corresponds to p then $\tilde{A}$ corresponds to ~p.

   iii. pv~p is a tautology. To it corresponds U.

   iv. ~(((p∧q)v(pv~q))v~p) is logically false. The empty set (IIIA3) corresponds to it.

2. There is also an analogue between connectives and Boolian operations. This is as shown.

   i. The correspondence is:

      a. ~p with $\tilde{P}$

      b. p∧q with $P \cap Q$

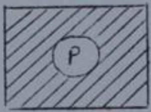
      c. pvq with $P \cup Q$

      d. p→q with $\widetilde{P \cap \tilde{Q}}$

      e. p↔q with $\widetilde{P \cup Q} \cup (P \cap Q)$

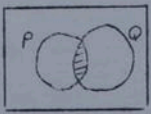
ii. Pictorially this looks like this.

a.



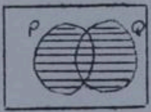
$\sim p$ true for shaded area, i.e., $\tilde{P}$

b.



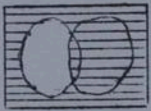
$p \wedge q$ true for $P \cap Q$

c.



$p \vee q$ true for $P \cup Q$

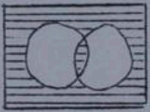
d.



$p \rightarrow q$ is true for

$\widetilde{P \cap Q}$

e.



$p \leftrightarrow q$ is true for

$\widetilde{P \cup Q} \cup (P \cap Q)$

iii. The result of the analogous operation in set theory is the set of situations where the new compound proposition is true.

C. Formalism in mathematics.

1. Types of statements.

i. Axiom - a statement which is accepted as a priori true. The need for these will be pointed out later.

ii. Definition - a statement which says what is meant by some term or symbol.

iii. Theorem - a statement which can be concluded (i.e., deduced) from axioms and definitions.

2. What is proof.

i. Deduction vs induction.

  a. Deduction is another word for "rewording". Deduction never brings anything new to a collection of statements, it just rewords statements to reveal the truths that were there to start with. In this sense, it is like a juice extractor--no more (and frequently less) juice comes out than was contained in the oranges put in.

  b. Induction is another word for "guessing." Induction brings new statements to a collection of statements. It is frequently difficult to be sure of the truth of induced statements. Most of the real results of the empirical sciences such as biology are the result of induction. You can't prove an induction, you can only justify it, or show why it constitutes a reasonable guess.

ii. Hence, we can only prove by deduction (i.e., a series of rewordings) to show that what we are trying to prove is just a rewording of other statements which we want to accept as true (be they axioms, definitions, or other theorems).

  a. Giving an example does not constitute a proof, although an example may have much heuristic value in an explanation.

  b. Giving 73 examples does not constitute a proof either.

iii. Duplation example when two wrongs make a right.

  a. 19 x 5

  9 x 10

  ~~4 x 20~~

  ~~2 x 40~~

  1 x 80

  95

  b. Always works but was not proved until modern times.

3. Conditions.

   i. Necessary.

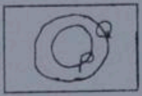
      a. p is necessary for q if $q \rightarrow p$.

      b. Pictorially, in the Boolian algebra analogue, p is necessary for q if the truth set for p contains the truth set for q.

   ii. Suffient

      a. p is sufficient for q if $p \rightarrow q$, (i.e., [q is necessary for p] $\leftrightarrow$ [p is sufficient for q]).

      b. Pictorially, p is sufficient for q if the truth set for q contains that for p.



   iii. To say p is necessary and sufficient for q is to say p is the same as q or $p \leftrightarrow q$.

4. Determination. A set of conditions is provided. This set of conditions determines the set of objects* which satisfy these conditions.

   i. Underdetermination.

      a. Occurs when the set determined by the conditions has more than one member.

      b. Example: {female, student at CSU, senior} underdetermines individual students, uniquely determines a set of students.

   ii. Over determination.

      a. Occurs when the set of objects satisfying the conditions given is empty.

---

*Object is a general term which also includes concepts, ideas, events, etc.

b. Example. The straight line passing through the points (-1,1), (0,1), (1-1).

iii. Unique determination.

a. Occurs when the satisfying set has one member.

b. Example. College Presidents living in Fort Collins.

V. <u>Analysis</u>.

A. Functions.

1. Definition two sets together with a rule which associates with each object in the first set a unique object in the second.

2. Example.

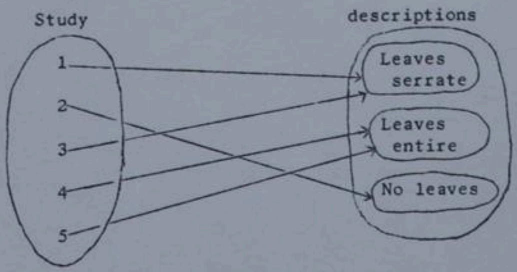
a. Let the first set be the feet in this room and the second be the shoes in this room, and the rule is "associate with each foot the shoe that it is wearing."

b. Both sets are the integers. The rule is "to each integer in the first set associate with it the integer in the second which is twice as large."

3. Functional relationships between sets of things will prove a very useful concept throughout this course.

VI. <u>Other mathematics.</u> There are many topics that we have not touched on in these three lectures. When you hear from me again later I will undoubtedly include discussions of other considerations as they arise. What we have discussed is a rather arbitrary and very small subset of what I could have chosen. What is important is not so much the content of these discussions but rather that something of a concept of what mathematics is be germinated.

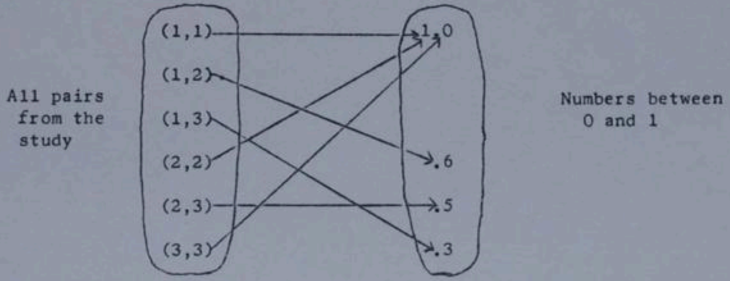
## CHARACTER 1 Leaf Margins



A CHARACTER is a function (cf p. 16) i.e. two sets together with a rule which associates with each member of the first set a unique member of the second. A character is a function whose first set is the study to be described and whose second set is a set of descriptions which describe different possible observable conditions of some basis for comparison. In this example there are 5 objects to be described; the basis for comparison is the leaf margin and the descriptions of this basis are leaves serrate, leaves entire, no leaves.

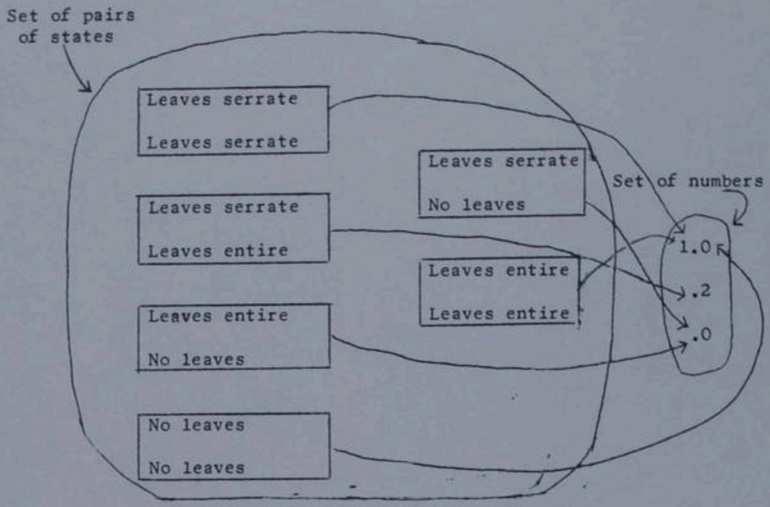
A similarity measure for a study to be classified is also a function whose first set is the set of all pairs of members from the study and whose second set is all the numbers from 0 to 1 inclusive. The function associates with each pair in the first set some number which indicates how mutually similar that pair is. For example:

If there were 3 members in the study, then the similarity measure might look like



In order to create a similarity measure from characters, it is necessary to associate with each character yet another function (sort of a preliminary or junior similarity measure). This function has for its first set all pairs of states in a character and for its second set the numbers 0 to 1 inclusive.

Continuing the first example



There are some restrictions on this function. A pair of identical states MUST be assigned 1. A pair of nonidentical states MUST NOT be assigned 1.

Let K(a) be the state of character K assigned to plant (a). Similarly, let K(b) be the state of character K assigned to plant (b). Let n(K(a), K(b)) be the number assigned by the above function to the state pair (K(a), K(b)).

The similarity measure will then assign to the plant pair (a, b) the number

$$S(a, b) = \frac{\sum_{\text{All characters } K} n(K(a), K(b))}{\text{Total number of characters}}$$

NO INFORMATION AVAILABLE

When this situation obtains for either plant a or plant b for some character K, the above formula is modified by setting n(K(a), K(b)) = 0 for that character and by reducing the denominator by 1.

Techniques of determing the "junior similarity measure" for a character.

1. Simple. All the states are very distinct and show little differentiating internal similarity. In this case this function is defined simply as n(k(a), K(b)) = 1 only if K(a) = K(b) otherwise 0.

2. Well ordered states. This occurs when there is some natural ordering for the states where proximity in the ordering indicates internal similarity between the states. For example, flower color might have five states:

red, light red, pink, pink-white, white

In this case red is more similar to light red than to pink etc.

An interpretive formula can be used to describe the "junior similarity measure", namely

$$n(K(a), K(b) = F(\triangle, c)$$

$$= \frac{2\ (c + 1 - \triangle)}{2c + 2 + \triangle c}$$

is the "distance" between states in the established ordering. For example, with the flower color above if

$K(a) = $ red and
$K(b) = $ light red

$\triangle = 1$, where as if

$K(a) = $ red and
$K(b) = $ pink

$\triangle = 2$ etc.

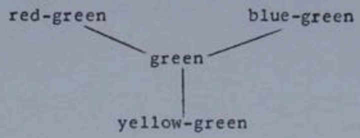
c is a constant which is your judgment of how far apart two states can be and still be partly similar. In the example if a $\triangle$ of 3 or greater is judged to indicate total dissimilarity between states then in the example states red and pink-white would be judged totally dissimilar. In this case

n(red, pink-white) = 0.

3. In the remaining instances, no well ordering of states is possible but an internal relatedness between certain states of the character in question is recognized. In this case the "junior similarity measure" must be represented in a matrix. For example, consider stem color: Assume the states are

green red green blue-green yellow-green

their internal affinities are viewed as



A matrix expressing this would be

| | green | red-green | blue-green | yellow-green |
|---|---|---|---|---|
| green | 1 | | | |
| red-green | .5 | 1 | | |
| blue-green | .5 | 0 | 1 | |
| yellow-green | .5 | 0 | 0 | 1 |

Here, for example:

n(green, red-green) = .5

while

n(red-green, blue-green) = 0.0

## CLUSTERING METHOD

The following discussion concerns itself largely with the activities of box 3 page 1.

In order to develop a mathematical methodology, it is important that the problem be formulated precisely, and that definitions of the concepts used to describe the problem be defined. From the discussions in this course so far, I am going to suggest that the following three generalizations about classification be considered. These three statements are virtually equivalent to the mathematical methodology, which we use (i.e., on which our computer programs are based, box 4 page 1).

### I. Generalizations About Clustering.

1. Convention: a classification for a collection of objects is a series of partitions for this collection, characterized by the equivalence relations $R_1$, $R_2$, ........,$R_M$, (see pages 7, 8, 9) with the property that if $j > k$ then $aR_k b \rightarrow aR_j b$.
2. A cluster of objects, to be a good cluster, should show some external discontinuity, i.e., should be isolated from the non-cluster members.
3. A good cluster should show internal continuity, i.e., should not break apart into isolated subclusters.

Clearly these generalizations are not completely true in all cases, nor do they embody the entire concept of phenotypic classification. However, they are relatively simple, not too difficult to accept tentatively, and they do lead to what has proven a relatively successful mathematical procedure for aid in classification.

### II. Definition of a Cluster.

In keeping with the generalizations, a cluster will be defined as follows.

A subset S, of the study, is said to be isolated for some value c,

$0 < c \leq 1$, if $p \epsilon S$, $q \notin S$ implies $S(p,q) < c$.

An acceptable cluster will be a subset of the study which is isolated for some similarity, c, but which contains no smaller subsets of the study which are isolated for $c' \leq c$.

Let S be an acceptable cluster for some study. Let S be isolated for the similarity value c. S will be called a c-cluster if $c' > c$ implies S contains smaller subsets which are isolated at c'.

Let $G_c$ be a relation for the study. $aG_c b \leftrightarrow S(a,b) \geq c$.

Let $R_c$ be a relation. $aR_cb$ if either $aG_cb$

or for some s, $aG_cs$, $sG_cb$

or for some s,s' $aG_cs$, $sG_cs'$, $s'G_cb$

or etc.

$R_c$ is an equivalence relation. (Page 8).

Show $R_c$ to be

Reflexive:

Symetric:

Transitive:

For any series of similarity values, $c_1, c_2, \ldots\ldots, c_M$, $i<j \rightarrow c_i > c_j$, the corresponding series of equivalence relations, $R_{c_1}, R_{c_2}, \ldots\ldots, R_{c_M}$, satisfies 1 (page 20).

That principles 2 and 3 are satisfied by the classes established by $R_{c_i}$ should also be clear.

## III. Method.

Where as there are virtually an infinite number of ways to choose decreasing series of similarity values, it is true that if there are N objects to be classified, there will be at most N distinct relations of the $R_c$ type, possibly fewer. This means that there are at most N ways to partition a study which is consistent with the 3 generalizations (on page 21.)

As a corollary to this, there are at most 2N-1 clusters which satisfy the definition of an acceptable cluster. It is not unreasonable to ask the computer to find them all.

## IV. Measures for Acceptable Clusters.

Of the set most 2N-1 clusters found for some study, some will be better than others. The following measures may be useful in helping to describe which of these suggested clusters are best suited for taxonomic purposes.

1. Moat of a cluster C. If C is a c-cluster, its moat is defined to be the similarity difference:

$$c \quad - \quad \underset{\substack{p \in C \\ q \notin C}}{\mathrm{Max}}[S(p,q)]$$

This number indicates the isolation of, or the amount of "empty space" around, C. When moat is high, it is not likely that members of C will be confused with non-members.

2. Connectedness. This is a measure of the internal "tightness" of a cluster. Connectedness is the number pair:

$$(K_t, K_a) \qquad \text{where}$$

$K_t$ is the total number of pairs in the cluster, $\frac{N(N-1)}{2}$ , and

$K_t$ is the actual number of pairs in the relation $G_c$.

## V. Graph Theoretic Interpretation.

1. A graph is a collection of points some pairs of which are connected. If all the pairs in the study which are in the relation $G_c$ (for some numerical value of c) are "connected" the result is a graph. A drawn representation of the study showing the connections is a good means of representing the results. This graph representation shows clusters as interlinked aggregates of points.

2. Articulation points. Let S be a c-cluster, and $a \varepsilon S$, a is an articulation point for S at c if the removal of a from the study results in the members of S belonging to more than one equivalence class under $R_c$. Articulation points may be intermediate forms for a study and may bring together otherwise distinct clusters. You may choose (and rightfully) to divide a c-cluster on the basis of the presence of an articulation point.