

AD

AD 629 29 DV

TR-1392

Part III

AUTOMATION OF THE ABC SYSTEM

Part III. Appendixes, Charts, and Illustrations

DDC
RECEIVED
OCT 2 1968
B



U S ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE
AND IS IN THE PUBLIC DOMAIN

U.S. GOVERNMENT PRINTING OFFICE
CLEARINGHOUSE

U.S. GOVERNMENT PRINTING OFFICE: 1968

421

Best Available Copy

AD

DA-2P020401A728
ANCMS Code: 5910.21.65071
HDL Proj: 01220

TR-1392

Part III

AUTOMATION OF THE ABC SYSTEM

Part III. Appendixes, Charts, and Illustrations.

by

Berthold Altmann

Walter A. Riessler

August 1968



U.S. ARMY MATERIEL COMMAND
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C. 20438

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE
AND SALE. ITS DISTRIBUTION IS UNLIMITED

CONTENTS

	Page	No.
ABSTRACT	3	3
ACKNOWLEDGMENT	4	4
PART I. LINGUISTIC PROBLEMS AND OUTLINE OF PROTOTYPE TEST		
A. Introduction	7	7
B. Linguistics	8	8
a. The linguistic Problem of Information Retrieval	8	8
b. Coordinate-Index Type Systems	10	10
c. The ABC - Method	12	12
d. General Considerations	13	13
e. Definitions	14	14
1. Terminology Used to Describe the Natural Language	14	14
2. Document vs Information Retrieval	16	16
f. Semantic Theory	16	16
g. The American Psychological Approach	18	18
h. Syntax and Semantics: The Thinking-Psychological Approach	19	19
i. The Thinking Machine	21	21
j. Discussion	22	22
k. Focalization upon Specific Requirements	24	24
C. The Prototype Automated ABC Retrieval System	24	24
a. Test Collection	25	25
b. Processing for Automation	25	25
c. The Measuring Tool and Its Applications	26	26
d. Statistical Problems	29	29
e. Elements Tested	30	30
f. Impact of the Number of Categories in a Given System	30	30
g. On-Line Retrieval	32	32
PART II. PROTOTYPE TEST (DESIGN AND ANALYSIS) AND PROCESSING FOR SECOND-GENERATION MODEL		
D. Model and Statistical Analysis of Automatic Prototype Test	33	33
a. Evaluation of the Results	33	33
b. The Recall-Relevance Relationship	36	36
c. The Standard Deviation	40	40
d. Decision about Relevance Formula	41	41
e. Smoothing Procedures	43	43
f. Reduced Evaluation Scales	45	45
g. Effect of the Number of Categories	46	46
h. Discrimination Power of a Category	47	47
i. Experiments for Improving D.	48	48
E. From Test Model to the Comprehensive Test and to the Second-Generation ABC System	49	49
a. The Categories and Their Applications	50	50
b. The Worksheet Approach	51	51
c. The ABC Retrieval Methods	53	53

F. Conclusions and Projections	55
POSTSCRIPT	60
REFERENCES	61
PART III. APPENDICES, CHARTS, AND ILLUSTRATIONS	
Appendix I --Computer Files and Programs	69
Appendix II-- The Second-Generation ABC Dictionary	73
Appendix III-- The Mechanical Standardization Process for ABC Descriptors.	75
Appendix IV-- Derivation of the Distribution Formula	77
Appendix V-- Participants in the Construction of the Test Collection.	79
Chart A Derivation of Test Collection and Queries.	80
Chart B Categories Used for Automated Test.	81
Chart C Dependence of D (Deficiency) on Progressive Acceptance of Decreasingly Relevant Documents as Relevant Ones.	82
Chart D Preliminary Worksheet	83
Chart E Worksheet for Structured Abstracts.	84
Chart F Thesaurus Automatically Derived from Input.	85
Chart G Flowchart for Automatic Standardization of Syntagmas (ABC Descriptors).	86
Chart H Filter Codes.	87
Chart I Selective Dissemination Worksheets.	88
Chart J Sample Page of Second-Generation ABC Dictionary	89
FIGURES	
1. Two-dimensional presentation of ranked order output	90
2. Normalized ranked-order output as presented in Figure 1	90
3. Relevance-recall curves derived from formula (1).	90
4. Recall-relevance curves as function of αK and πK	91
5. K as function of D.	92
6. Integral distribution of the D_i 's obtained in 3 test runs	93
7. Normalized standard deviation of D plotted vs the corresponding D of several test runs.	94
8. D_i 's of the 50 queries vs numbers of responsive documents for one test run	95
9. Frequency of evaluation numbers used by different evaluators (sample)	96
10. First smoothing method applied to the vectors of one evaluator.	97
11. Effect of reducing the 10-valued to a 2-valued scale upon the deficiency.	98
12. Effect of number of evaluation grades upon D.	99
13. D vs number of applied categories	100
14. The effectiveness of the individual categories vs their frequency within the document vectors	101
15. D_c (change of D by dropping category c) vs W_c (efficiency of category c).	102
16. Changes of D when the retrieval formula is modified by weight factors W_c	103

APPENDIX I
COMPUTER FILES AND PROGRAMS

It is obvious that the numerous mechanical retrieval runs performed during the test of the prototype automated system required a large amount of programming effort. Because this effort, valuable as it is, will not contribute to the day to day operation of the second-generation system. It has been omitted from the following list which contains only such computer files and programs as will assure the effectiveness and efficiency of the ABC storage and retrieval method. The different programs written in COBOL are as elements of one integrated system ready to be compiled and to perform all required operations and produce all required products in one continuous process⁵⁷.

The current computer programs completed or nearing completion produce and maintain the following magnetic tape files:

1. the ABC descriptors in their unrotated form by alphabetical accession and identification code
2. the ABC descriptors alphabetically sequenced by rotated keywords as a result of KWIC-program processing
3. the dated ABC descriptors withdrawn from the active file by subject terminology or subject-date combination and in unrotated form as File No. 1
4. the ABC descriptors related to selected subject categories in the format of File No. 2
5. the complete bibliographic information (subdivisions by fields, personal author, corporate author, title, contract or project number, year, security classification, each one directly accessible) for the active library holdings ; its primary organization by type of publication: reports, books, periodicals and periodical articles, etc.; and its secondary organization by shelf or accession numbers
6. the inventory of all bound book and periodical volumes by shelf number
7. the title information withdrawn from the active File No. 5
8. authority list of corporate author names and name variations organized by identifying code
9. index to periodical titles organized by the procurement source

10. category names arranged by category codes
11. asterisk terms and descriptor ABC codes - an index referring to the full-length ABC Individual descriptors as well as to the catalog entries of corresponding titles
12. SDI printouts: user profiles accept: (a) a combination of **one** specified category and of (b) keywords in one particular category dictionary

The programs listed above generate the following printouts:

1. lists of unrotated ABC descriptors arranged by codes and indirectly by date
2. alphabetical index to terms contained in hyphenated phrases
3. ABC descriptor dictionaries in rotated format in alphabetical sequence by keyword, for entire collection, for categories, and for retired materials
4. accessions-bulletins of two-column printout with category headings inserted from File 10 above
5. complete sets of catalog cards ready for filing into personal author, corporate author, title, subject, parameter, ABC-subject, contract or project number, AD-number, etc., catalogs and shelf lists
6. rotated title list (one-line KWIC program)
7. cumulative document title catalog in book form by accession or bulletin number
8. cumulative rotated title listing for books, reports, etc. in KWIC format
9. lists of periodical holdings
10. periodical title lists by renewal date and vendors' names
11. punched cards for all issues of subscribed periodicals
12. lists of titles withdrawn from current files
13. catalogs for historical files

14. catalogs of all titles requiring downgrading of security classification
15. corporate author authority lists in 3 x 5 card format
16. corporate author cross references on 3 x 5 cards to be filed in catalog
17. periodical vendor lists by vendor code as well as by vendor name
18. texts of subject categories used as headings in accession bulletins and as titles of category ABC dictionaries
19. title catalogs arranged by ABC descriptor code
20. title catalogs arranged by shelf number
21. lists of periodical issues not received (to be mailed to vendor)
22. frequency counts of documents broken down by ABC code within each category
23. lists of keywords by categories with frequency counts of related documents and descriptors (documents as well as descriptors are identified)

The category term dictionaries and their frequency counts will be used (a) to assist the professional man in entering the system from a terminal provided with teletype or electronic display reception equipment; (b) to develop vector numbers by a mechanical process that is to replace the psychometric evaluations and to organize the content of the collection under the assumption that this preparation and integration of information establishes multiple links to subjects of interest and prepares a spatial distribution or classification for effective mechanical retrieval; (c) to produce continuously changing measures of relatedness as well as of significance, to automate step by step the development of and the assignment to categories; and (d) to standardize the descriptor phrases, their linguistic components and their syntax by performing as a computerized effort the editorial changes of the analytical statements (ABC descriptors) through comparison with the information stored in a variety of formats, by categories, types of associations and terms and phrases with frequency counts so that human effort ultimately will be required only to supplement what is missing in current listings or stored tables.

In addition to the careful screening process and the organization of the selected information by a system of overlapping categories for more efficient retrieval operations, the filter code discussed in previous reports (see chart 1) will make it possible to pin point the type of publication, the particular phase of the development, and the degree of difficulty a particular investigator is interested in studying. Such an organization of the stored references will help eliminate the retrieval of unwanted materials in the manual or automatic retrieval process.

APPENDIX 11
SECOND-GENERATION ABC DICTIONARY

Of the numerous computer programs that have been developed the one to generate the second-generation ABC dictionary deserves particular attention.

The program is operational¹⁵. A sample of the dictionary is shown as chart J. The following characteristics of the dictionary are worth noting:

1. The legibility is improved by eliminating the throw-back. Instead, overflow is printed on the subsequent line rather than on the same line. Markers clearly indicate the start of each descriptor.
2. The text of the descriptor following the keyword in the index window is alphabetized for 30 digits to organize the ABC descriptors within the keyword cluster by use of standardized connectors such as prepositions, participles, and infinitives.
3. The maximum length of the descriptors is increased to 450 characters to improve the descriptions and to forego the hyphenated phrases that were necessary for brevity with the first generation format.
4. When any term fails to make a substantial contribution to the meaning of a given descriptor that descriptor is not assigned to the cluster for the deficient term. This restriction depending on the decision of the analysts is essential in confining approaches to useful information and is utilized in addition to the stop list of permanently blocked words.
5. Of the other terms that are also eliminated to streamline and compact the descriptors we list: (a) the parameters that form a separate title index (by alphabetized names of materials, components, devices, etc., subdivided by parameters and sub-subdivided by the numerical values in ascending order); and (b) the filter codes (Chart H) denoting the types of publications, and which are used to organize the title index of the ABC subject dictionary.
6. The size of the dictionary is further reduced by the use of hyphenated phrases. Nevertheless, access to the not-alphabetized (and therefore not-clusterforming) component words within these phrases can be provided through a mechanically produced index of cross-references.

7. About 40 to 50 ABC dictionaries in different slightly overlapping categories will be issued. This is a by-product of the vector analysis performed (pages 50 to 51) for automated retrieval operations. (The analysts relate each document to 10 to 15 of about 190 categories. Compilation and printing of the ABC descriptor dictionaries and the ABC term dictionaries for the entire collection and important categories, the accession bulletins, current awareness listings, various forms of card and book catalogs are automated. The retrieval operations, however, can be performed manually.*

* This semi-automatic method might be preferable for smaller organizations without computers of their own, especially if they have some access to a computer center.

APPENDIX III
THE MECHANICAL STANDARDIZATION PROCESS FOR ABC DESCRIPTORS

To illustrate the mechanical standardization process we consider the following one-sentence abstract of a paper. "A parallel spiral antenna system with immunity to nuclear-blast gamma radiation was designed for a chirp-radar fuze in a Nike-X antimissile missile." In the process of transforming this sentence into an ABC descriptor, the verb phrase "was designed" is replaced by the noun "design," so that without change of meaning the ABC descriptor will read: "Design of a parallel spiral antenna system with immunity, etc."

To standardize this or any other ABC descriptor and to produce a structured description, two factors are of pre-eminent importance. The logical sequence of the component ideas and their verbal expressions, and the consistency of connectors (prepositions, participles, infinitives, etc.) which combine the component ideas with each other and produce the conceptual entity we call an ABC structured abstract.

The HDL method is implemented by providing a questionnaire for the research analyst. The questions must be broad enough to cover every subject detail encountered and they must be arranged in a sequence in which concepts are strung together to form logical unambiguous ABC descriptions. Moreover, each individual question must relate to one standardized connector in order that the computer may produce structured (ABC) descriptions from the input of the encoded answers which the analysts have recorded on the questionnaire. In filling out the questionnaire the analyst will consider the questions merely as constant reminders to cover all essential aspects that a given document may contain; but he will remain completely unaware that the questions arranged by letters or numbers will encode his answers, and that the code will not only determine the order, but also the connector the computer will use in compiling, completing and printing the English language description. The analyst will also not be concerned with such details as to whether some questions must precede, and others follow, the main subject of the description in order to generate the standardized sequence of the component elements.

For a very general understanding of the properties and the potential of structured-abstract approach we present on chart D the still experimental questionnaire currently used by our analysts.

Furthermore, the manual will contain references that lead the analyst to the particular category-term dictionary in which pertinent standardized terminology has been organized by computer.

The actual questionnaire form which the research analyst will interpret is a matrix with a number of columns, to assure unambiguous answers. On chart E we illustrate the worksheet completed by an analyst who processed the subject content of the paper listed above. The first column is used to record all the answers relating directly to the main subject: "antenna"; the second column (2a) to record the answers related to (or modifying) the term "fuze"; and the third column (2b) provides an answer that modifies the term "radiation"; and the last column (3) the word "missile."

The governing subject of the main set as well as of the subsets is indicated by a subsequent colon. All terms and phrases in each set or subset (divided by hyphens) are directly and exclusively related to the respective governing subject.

When placed into sequence and combined by standardized connector terminology the computer will produce an ABC structured abstract that reads: "Design study of parallel spiral antenna: in fuze, resistant to gamma radiation fuze: in anti-missile, using Chirp radar radiation: produced by nuclear blast missile: Nike-X."*

To eliminate any misunderstanding we repeat that in this example both phrases: "in anti-missile" and "using Chirp radar" are modifiers of the term "fuze."

* The described process has been computerized and tested.

TEXT NOT REPRODUCIBLE

APPENDIX IV DERIVATION OF THE DISTRIBUTION FORMULA

To acquire a ranking function that gives the distribution of the relevant documents of set B among the non-relevant set A, we divided the procedure into $m+1$ steps in the following way. We ranked first B within the subset A_1 of A. Second, we repeated this process with A_1 in subset A_2 of A and so on. Finally we ranked A_m within A and completed a ranking of B in A. For each step we used the same ranking function $f(x)$. The location of a document originally located at X will be after the first ranking procedure:

$$x_1 = f(X) = f_1(x)$$

After the i^{th} ranking its location will be at

$$x_{i-1} = f(x_{i-1}) = f_1(x)$$

f_{n+1} can be derived from f_n through the operation

$$f_{n+1}(x) = f_n(f(x))$$

We assume that $f(x)$ deviated from $f(x) = x$ only slightly and therefore write

$$f(x) = x - \epsilon g(x)$$

where

$$g(0) = g(1) = 0 \quad (1)$$

For a sufficiently small ϵ we can write:

$$f_{n+1}(x) = f_n(x - \epsilon g(x)) = f_n(x) - \frac{\partial f_n}{\partial x} \epsilon g(x)$$

or

$$f_{n+1}(x) - f_n(x) = -\frac{\partial f_n}{\partial x} \epsilon g(x) \quad (2)$$

If n is considered a high number and ϵ a small one so that $\epsilon n = \gamma$ is finite we can transform the left side of equation (2) into the differential ratio

$$\frac{\partial f}{\partial \epsilon} = -g(x) \frac{\partial f}{\partial x}$$

For $g(x)$ we substitute the function $x(1-x)$ (which may be the simplest way to fulfill the condition of equation (1)) and arrive at

$$f = F(-y + \log \frac{x}{1-x})$$

The still unknown function F can be calculated because whenever $y = 0$ then $F(x) = x$. The result is

$$F(y) = \frac{1}{1+e^{-y}}$$

and therefore

$$f(v_1 x) = y = \frac{1}{1 + \exp(y - \log \frac{x}{1-x})}$$

an equivalent of

$$\left(\frac{1}{x} - 1\right)e^{\frac{y}{2}} = \left(\frac{1}{y} - 1\right)e^{\frac{-x}{2}} \quad (3)$$

Formula (3) shows a symmetry in x and y , the value of y can be replaced by formula (3) on page 35.

The problem of the distribution of the relevant items among the non-relevant ones in a given collection can also be treated as a diffusion problem where only the product t of both the diffusion coefficient and the diffusion time is the important element. The distribution functions are more complex, but resemble those developed above. We have preferred our method because of its mathematical simplicity. For high vs , corresponding to small Ds we have the relationship of our D and t approximated by the formula:

$$D = \frac{2}{\sqrt{\pi}} \sqrt{t} \quad \text{where } \pi = 3.1416 \dots$$

**APPENDIX V
PARTICIPANTS IN THE CONSTRUCTION OF THE TEST COLLECTION**

Baba, A. J. (BS)
 Boykin, C. (BS)
 Cohen, M. M. (PhD)
 Colbert, R. (BS)
 D'Angona, D. S. (BS)
 Friedberg, I. S. (MS)
 Gibson, H. F. (MS)
 Goldfarb, R. (BS)
 Hine, M. K. (MA)
 Isler, W. E. (MS)
 Jantz, R. (BS)
 Manion, F. M. (MS)
 Marsh, D. S. (BS)
 Mary, D. J. (BS)
 McCall, T. D. (BS)
 Mesrobian, A. (BS)
 Meyler, O. L. (BS)
 Miller, J. (BS)
 North, G. D. (BS)
 Patterson, M. S. (BS)
 Redcay, P. W. (none)
 Riessler, W. A. (PhD)
 Sommer, H. (PhD)
 Soper, W. L. (MS)
 Tucker, R. W. (BS)
 Tuttle, J. E. B. (BS)
 Wall, R. E., ARC (PhD)
 Watkins, S. (BS)
 Williams, W. K. (BS)

PhD	4
MA or MS	6
BS	18
none	<u>1</u>
	29

Chart A. Derivation of test collection and queries.

	<u>A. Test Collection</u>	<u>B. Queries</u>	<u>C. Retrieval Process</u>	<u>D. Evaluation Tool</u>
Manual Test:	3600 Documents	139	Manual (4 Different Groups of Operators)	Relevance and Recall
Automated Test:	300 Documents	50	Automatic (Vector Method Superimposed)	Deficiency with respect to Perfect Ranking

Chart B. Categories used for automated test.

1) characteristics, parameters, data	31) oscillation
2) theory, analysis	32) modulation, demodulation
3) Design, development	33) transmission, communication
4) electric, electromagnetic	34) detection
5) magnetic, ferrite	35) discrimination
6) acoustic, hypersonic, ultrasonic	36) synchronization, tuning, phase
7) optical, light	37) stabilization
8) photoelectric	38) automatic control
9) magnetomechanical, piezo	39) simulation, analog
10) thermal, temperature	40) switching, logic
11) tunneling (tunnel effect)	41) memory, storage
12) junction (space charge)	42) computer
13) field effect	43) radar, sonar
14) dielectric, ferroelectric	44) maser, laser
15) parametric (varactors, ...)	45) space science
16) linear devices and effects, (resistors, capacitors, ...)	46) military, weapons
17) diode, rectifier	47) manufacturing
18) transistor	48) miniaturization
19) non-linear devices and effects exc. nos. 17 and 18	49) thin-film
20) functional units	50) measuring, testing, observation
21) circuit	51) improvement
22) instruments, equipment, appliances (ready for use)	52) germanium, silicon, selenium
23) systems	53) compound semiconductors
24) noise	54) non-semiconducting material
25) interference, interaction	55) power, energy
26) reliability, vulnerability, aging, failure	56) pulse
27) isolation, shielding	57) frequency
28) generation of any kind	58) radio waves
29) conversion of any kind (energy, frequency, signal, ...) transducer	59) microwaves, millimeter-waves
30) amplification, gain	

Chart C. Dependence of D (deficiency) on progressive acceptance of decreasingly relevant documents as relevant ones.

NUMBER OF CLASSES INCLUDED	$r_i \times S_i$	D in %	ERROR INTERVAL %
1	1196	1.5	+ - 2.0
1 through 2	12532	1.7	+ - 0.7
1 through 3	31628	3.2	+ - 0.8
1 through 4	44641	3.5	+ - 0.8
1 through 5	54423	4.3	+ - 0.9
1 through 6	70446	3.8	+ - 0.7
1 through 7	85141	3.9	+ - 0.6
1 through 8	92906	4.4	+ - 0.7
1 through 9	96611	4.7	+ - 0.7

Chart D. Preliminary worksheet.

Code Questions

What is the main subject?

A	Form of Publication? Work phase? Type of effort?	A of
B	Which are the properties, characteristics?	adjective phrase
C	What is the shape, form?	adjective phrase
D	What is the physical phase?	adjective phrase
E	How is it produced, caused? (Tool, method)	produced by E
F	What is it influenced by or changed by?	influenced by F
G	What is it related to?	related to G
H	What is it a component part of?	being part of H
I	What is it limited to?	limited to I
Ia	What is excluded?	without Ia
J	What is it designated by?	designated J
K	What is it simulated by?	simulated by K
L	What is it modelled by?	modelled by L
M	What materials does it consist of? or what materials is it related to?	of M
N	What components does it consist of?	of N
O	What devices does it consist of?	of O
P	What instruments does it consist of?	of P
Q	What is its purpose? What is it made for?	for Q
R	What is it resistant to?	resistant to R
S	What is it vulnerable to?	vulnerable to S
T	What does it cause or effect?	resulting in T
U	On what does it have influence	of U
V	What is it performing or operating?	participle
W	What principle, energy, or instrument is applied?	using W
X	What is the reason or cause?	because of X
Y	What is it similar to?	like Y
Z	What is its environment?	in or at Z
AA	What is the location in space? (Table)	"in" "at" AA
AB	What is the location in time? (Table)	"during" AB

GRAPHIC NOT REPRODUCIBLE

Chart E. Worksheet for structured abstracts.

DOCUMENT CODE: A
 SHELF NUMBER: 14-302

LINE	COLUMN 1 REFERS TO:	COLUMN 2 REFERS TO: 1F	COLUMN 3 REFERS TO: 1G	COLUMN 4 REFERS TO: 2F
A	fusion study			
B	paraffin, alpha			
C	antenna	fuze	gamma radiation	anti-missile missile
D	fuze	anti-missile missile	nuclear blast	
E				Mike-X
F				
G				
H				
I				
J				
K	gamma radiation			
L				
M				
N		chirp-radar		
O				
P				
Q				
R				
S				
T				
U				
V				
W				
X				
Y				
Z				

*Main subject THE FOLLOWING CARDS ARE PUNCHED FROM THE WORK SHEET.

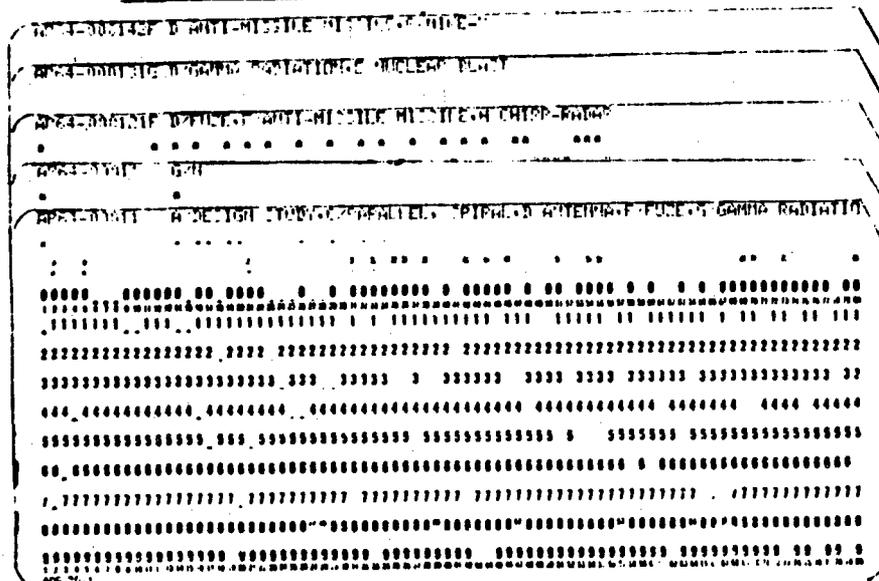


Chart E. Thesaurus automatically derived from input. (short sample)

<u>A</u> <u>Terms</u>	<u>Modified by</u>	<u>Associative</u> <u>Code</u>	<u>ABC</u> <u>Code</u>
Antenna	Design	A	CFA
	Fuze	H	CFA
	Gamma Radiation	S	CFA
	Parallel	C	CFA
	Spiral	C	CFA
Fuze	Anti-Missile Missile	H	CFA
	Chirp Radar	W	CFA
Gamma Radiation	Nuclear Blast	E	CFA
Missile	Nike-X	J	CFA

B

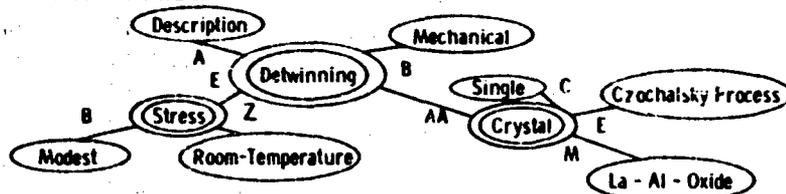
<u>Terms</u>	<u>Used as modifiers of</u>		
Anti-Missile Missile	Fuze	H	CFA
Chirp Radar	Fuze	W	CFA
Design	Antenna	A	CFA
Fuze	Antenna	H	CFA
Gamma Radiation	Antenna	S	CFA
Nike-X	Missile	J	CFA
Nuclear Blast	Gamma Radiation	E	CFA
Parallel	Antenna	C	CFA
Spiral	Antenna	C	CFA

Chart G. Flowchart for automatic standardization of syntagmas (ABC descriptors).

I. ORIGINAL ABSTRACT

Crystals produced by the Czochalsky process frequently show twin boundaries. An experiment was conducted by which twins in Lanthanum - Aluminum - Oxide were removed by the application of modest stress at room temperature. The paper describes the detwinning method.

A. Tree Presentation



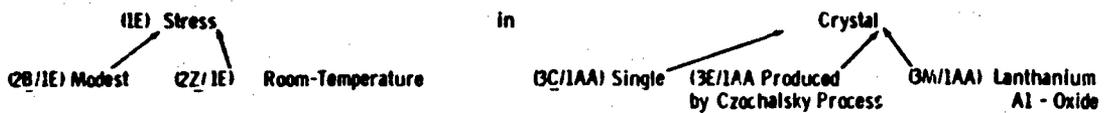
II. CONTENT ANALYSIS

B. Format of Computer-Stored Analysis (taken from worksheet)

1A	Description				
1B	Mechanical	2B	Modest	3C	Single
1I*	Detwinning	2/1E*	Stress	3/1AA*	Crystal
1E	Stress			3E	Czochalsky
		2Z	Room Temperature	3M	La - Al - Oxide
1AA	Crystal				

III. COMPUTER PROCESSING METHOD (transformation of II.B.)

Step 1: Introduction of Standard Connectors and Preliminary Arrangement:
Description of Mechanical Detwinning produced by



Step 2: Intermediary Process: Modifiers and their standard connectors are properly sequenced, those coded A-E in front of, all others behind the term that is modified.

Step 3: Standardized Computer Product: "Description of mechanical detwinning produced by modest stress at room temperature", in single crystal produced by Czochalsky Process of Lanthanum - Aluminum - Oxide".

* The comma indicates that the subsequent phrase is directly related to the main subject ("detwinning"), but its component elements refer to the main subject of this particular phase ("crystal").

Chart II. FETTER codes.

A	State-of-the-art surveys
B	Bibliography, abstracts
C	Collections, proceedings
D	Design, development and engineering studies and reports
E	Elementary, popular, introductory studies
F	Feasibility studies
G	Graphs and tables
H	Historical studies
I	
J	Dictionaries, lexicons
K	Computer programs and simulation
L	
M	Mathematical and statistical studies
N	
O	
P	Production engineering
Q	
R	Research: applied and theoretical
S	Standards, specifications
T	Tests: laboratory
U	Tests: field
V	Test equipment and procedures

Chart J. Sample page of second-generation ABC dictionary.

K E Y W O R D S	C O D E
DIELECTRIC COATING •	WYJ
SLAB •	SMH
MAGNETOSTATIC FIELD •	SCN
PRESENCE OF MAGNETIC-FIELD •	AASZ
•••MODAL ANALYSIS AND SYNTHESIS OF RADIATION AND DIFFRACTION OF ELECTROMAGNETIC-WAVE IN COLLISIONLESS PLASMA •	XAG
•••EFFECT OF NUCLEAR-WEAPON BURST ON ELECTROMAGNETIC-WAVE PROPAGATION •	TEA
•••USSR COMPILATION OF ABSTRACT ON ELECTROMAGNETIC-WAVE PROPAGATION AND UPPER-ATMOSPHERE •	TNL
CONSIDERING ELECTRON VELOCITY AND COLLISION •	WIP
ATMOSPHERE FROM NUCLEAR-WEAPON DETONATION •	SGH
PERIODIC-STRUCTURE EXCITED BY A WAVEGUIDE •	TND
TAPS-TSP12 •	VFA
TURBULENT-ATMOSPHERE •	SSA
MAGNETOSTATIC-WAVE IN FERROMAGNETIC POWER-LIMITER •	SIN
PARABOLIC PLASMA SLAB AT ARBITRARY-INCIDENCE-ANGLE •	WAA
•••EFFECT OF HIGH ALTITUDE NUCLEAR-BURST ON ELECTROMAGNETIC-WAVE TRANSMISSION THROUGH WITH AN EXPONENTIALLY-INCREASING CHARGE-DENSITY •	ABBF
LOWER-ATMOSPHERE •	MSV
STRATIFIED PERIODIC-DIELECTRIC-MEDIUM •	TTN
•••TEMPORAL AND SPATIAL DISPERSION OF ELECTROMAGNETIC-WAVE-PROPAGATION IN SQUARE-WAVE-AND COATED INFINITE CYLINDERS •	SGH
•••FLANNING GUIDE FOR SATELLITE ECHO SUBSYSTEM AGAINST ENEMY ELECTROMECHANICAL-INSTRUMENTS •	WRY
LITERATURE •	NJY
•••ANNUAL REPORT ON ELECTROMETER AND ACCESSORY •	ZNG
•••EFFECT ON UNIJUNCTION TRANSISTORS OF EPJIS ELECTRON •	SMK
•••VISIBLE AND UV RADIATION FROM METAL SCORPED BY ELECTRON •	SEK
INSTRUMENTATION •	UDR
ELECTRON BEAM •	AAGV
•••ELECTRON ATTENUATION IN GOLD FILM FOR SLOW ELECTRON BEAM •	ABMY
•••COUPLED-MODE THEORY OF INTERACTION BETWEEN A ROTATING ELECTRON BEAM AND A WAVEGUIDE •	HUD
	MOE
	ZYP
	UBQ
	ZMF
	ZLH
	ZLN
	ACDB

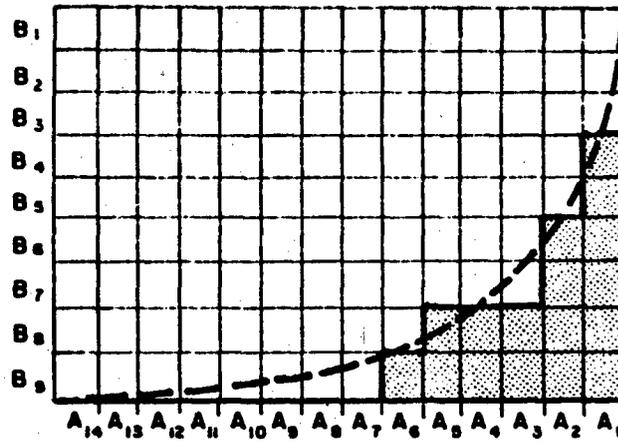


Figure 1

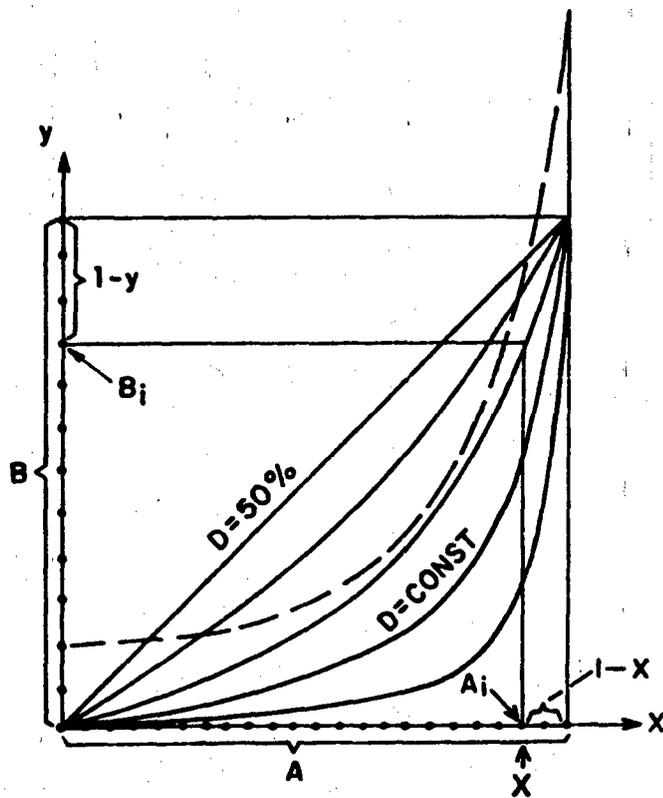


Figure 2

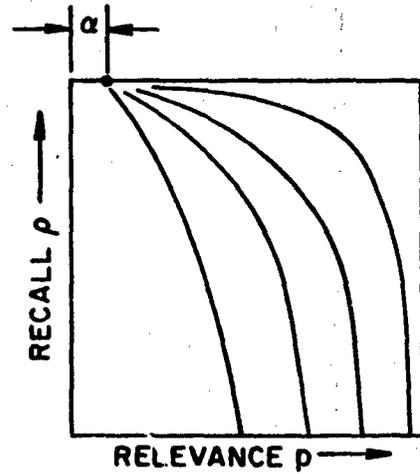


Figure 3

- Figure 1. Two dimensional presentation of ranked order output.
 Figure 2. Normalized ranked-order output as presented in figure 1.
 Figure 3. Relevance-recall curves derived from formula (1).

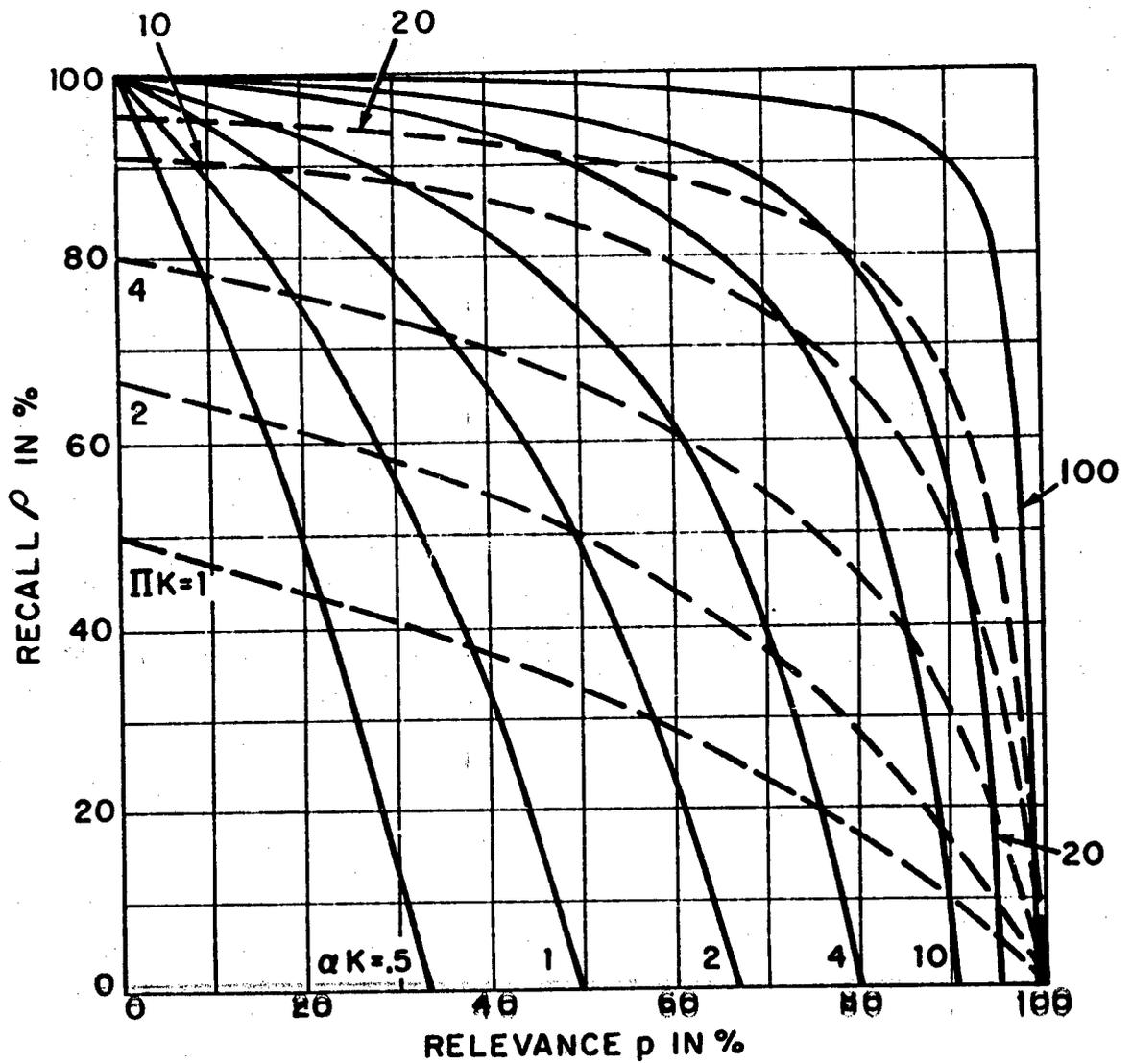


Figure 4. Recall-relevance curves as functions of αK and πK . (ρ is the relevance of the entire collection with respect to a particular query and π is the ratio of the documents withdrawn)

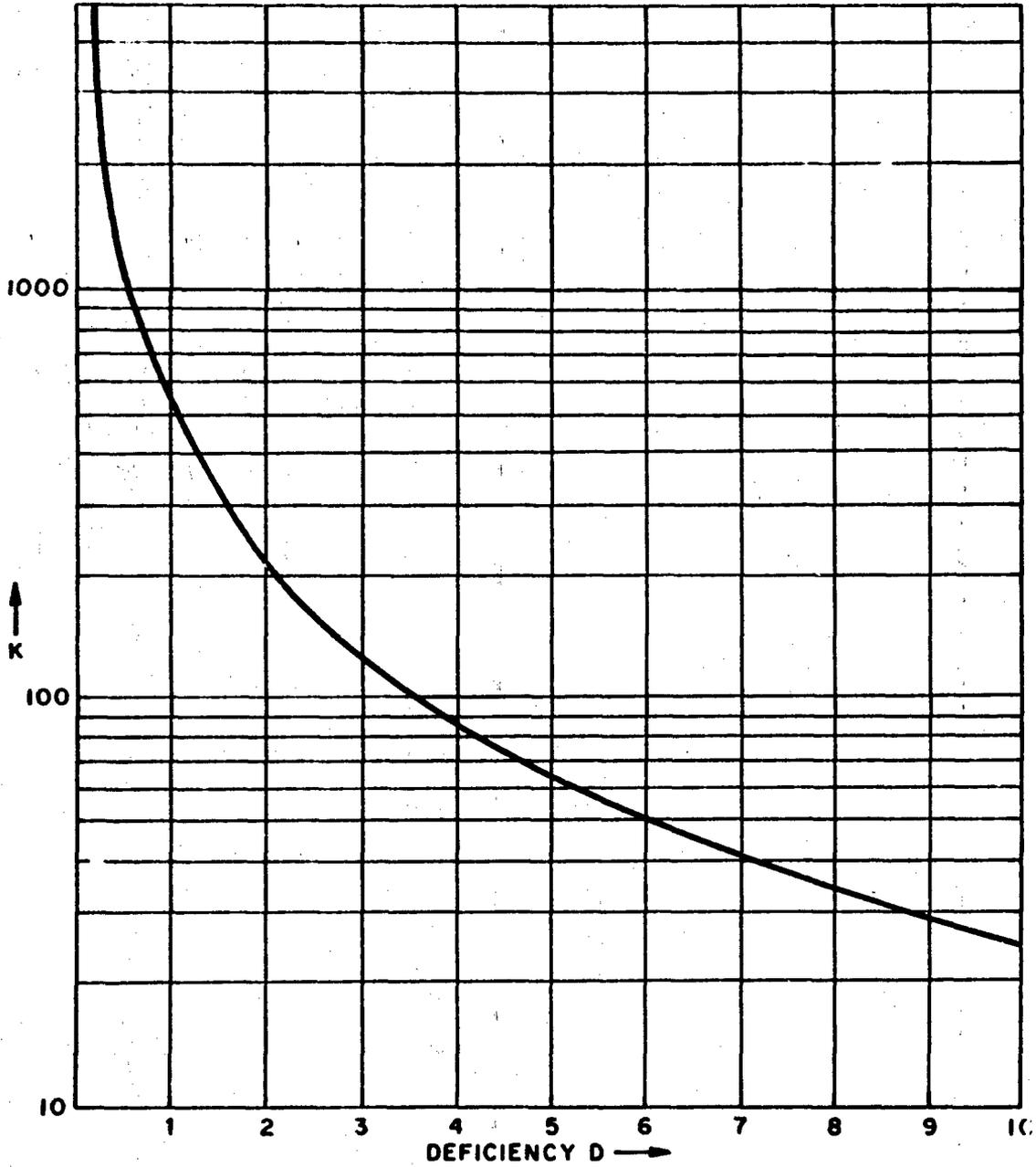


Figure 5. K as function of D.

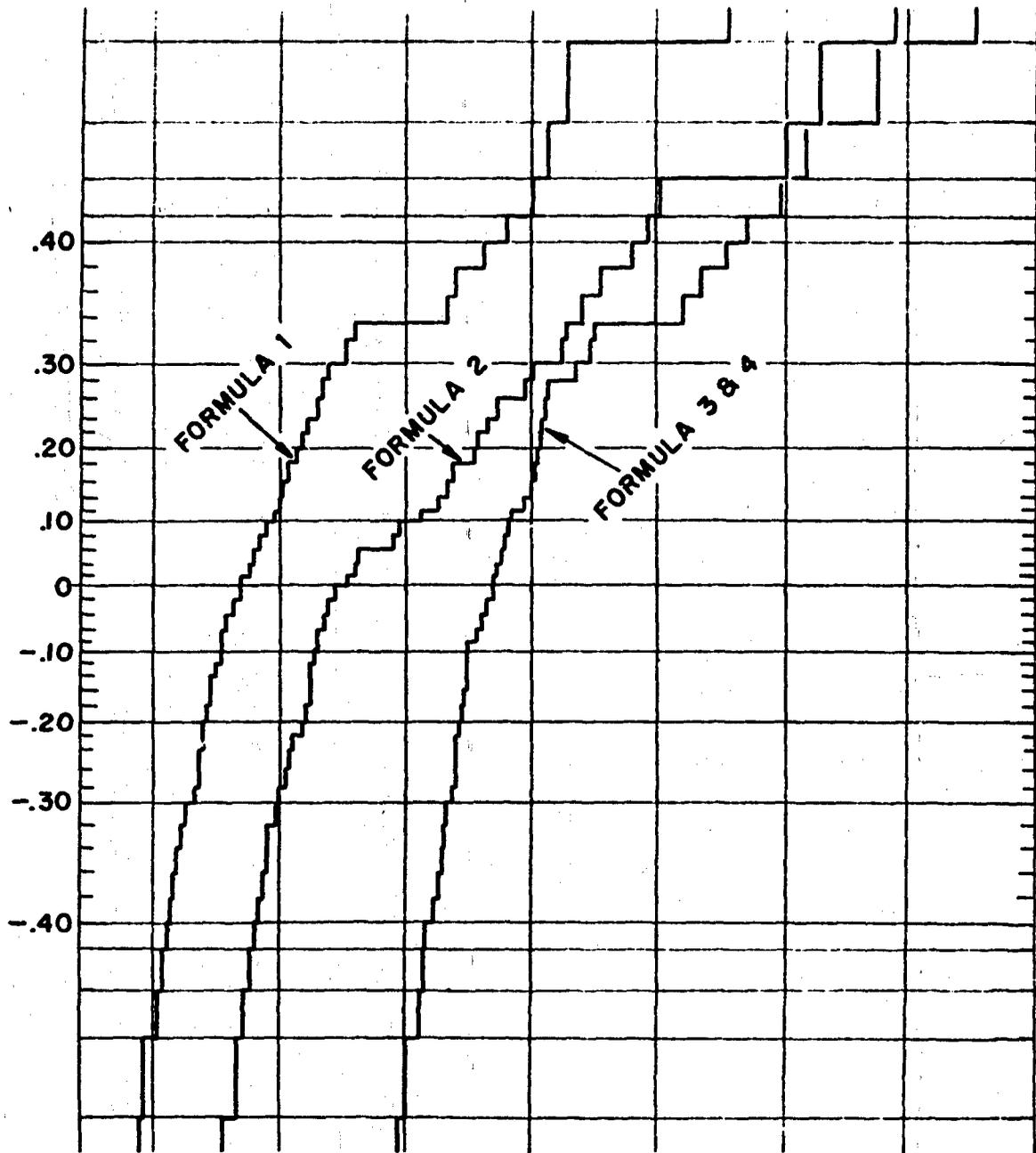


Figure 6. Integral distribution of the D_1 's obtained in 3 test runs.
 (a Gaussian distribution would appear as a straight line
 in this presentation)

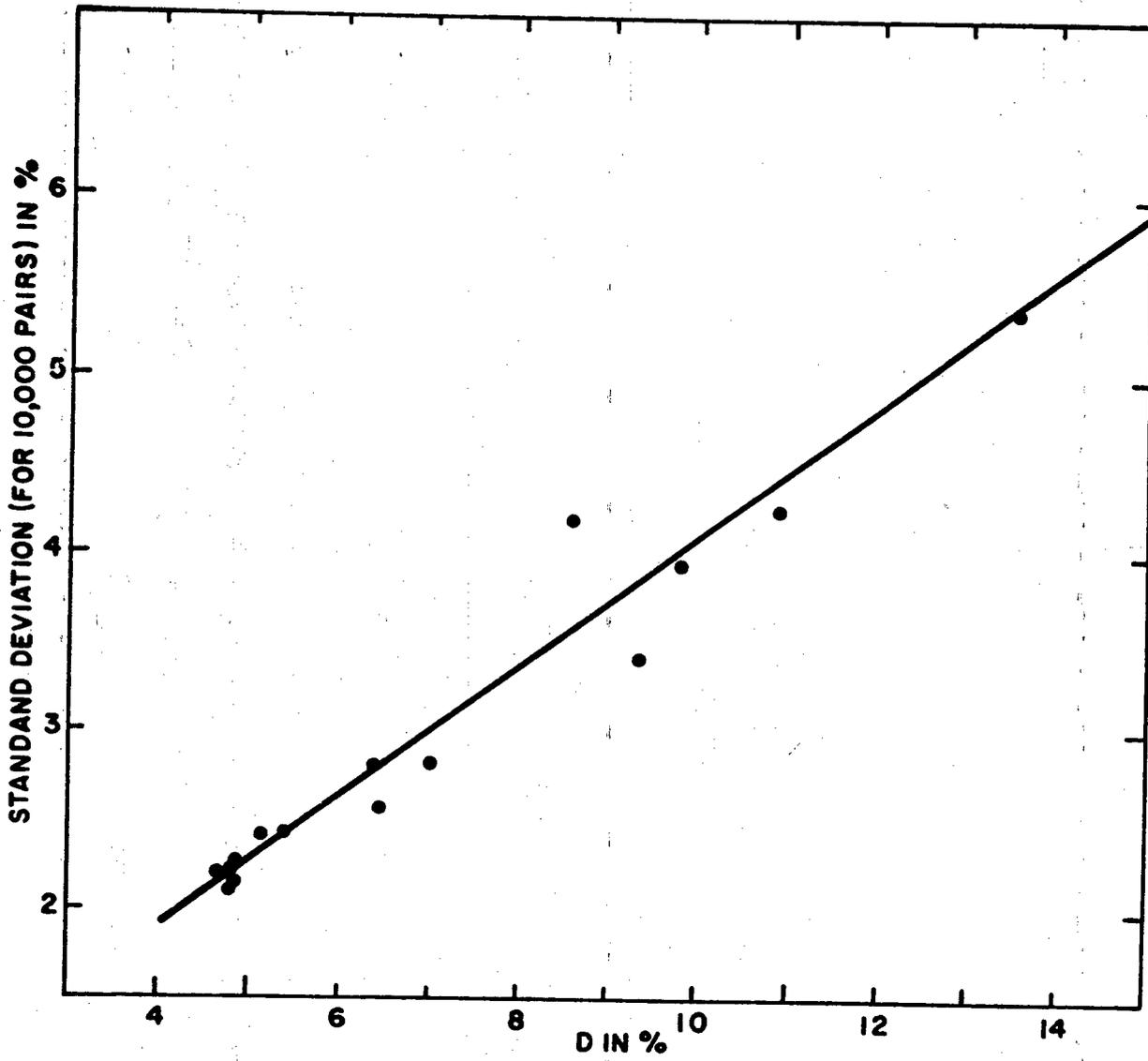


Figure 7. Normalized standard deviation of D plotted vs the corresponding D of several test runs.

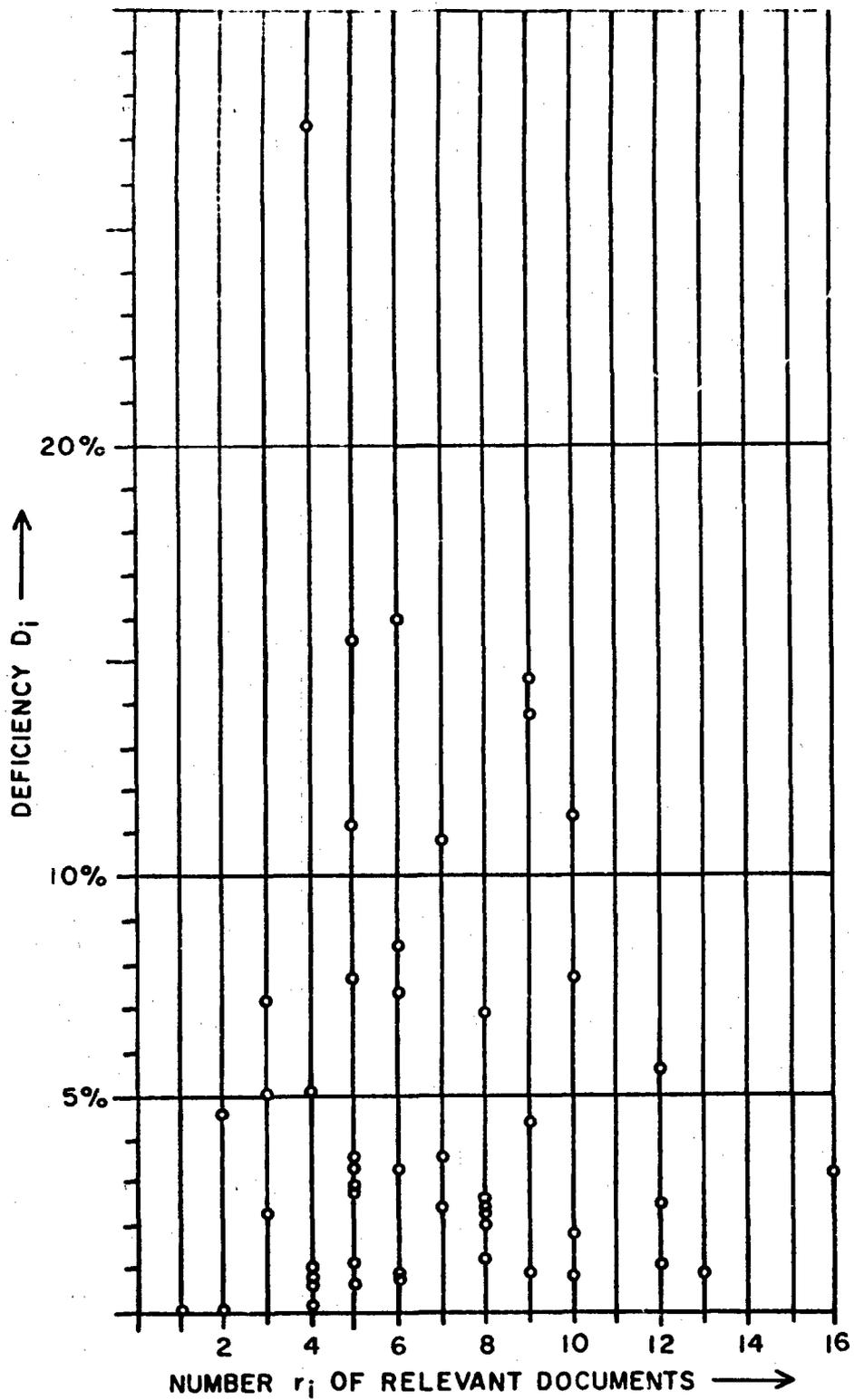


Figure 8. D_i 's of the 50 queries vs numbers of responsive documents (for one test run).

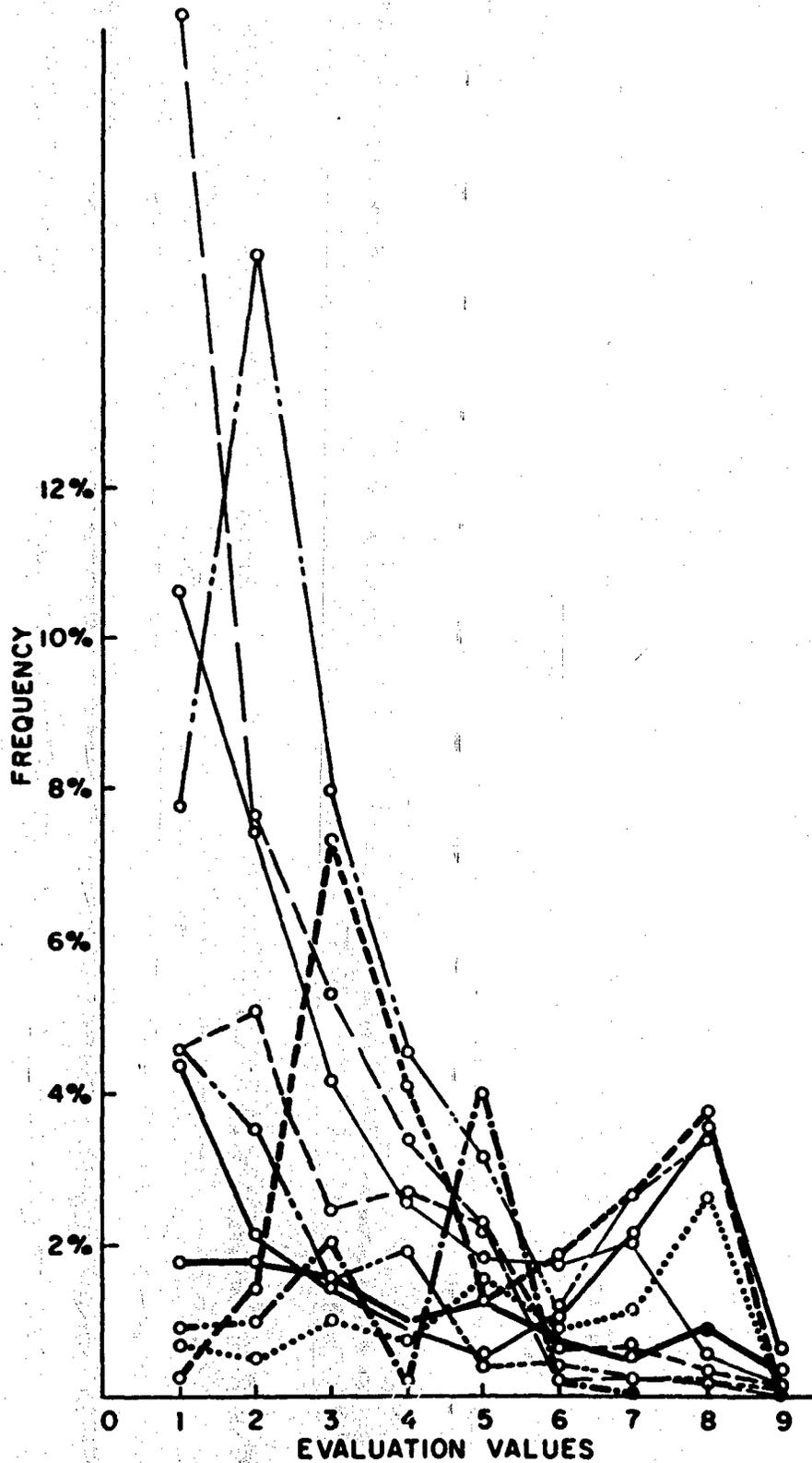
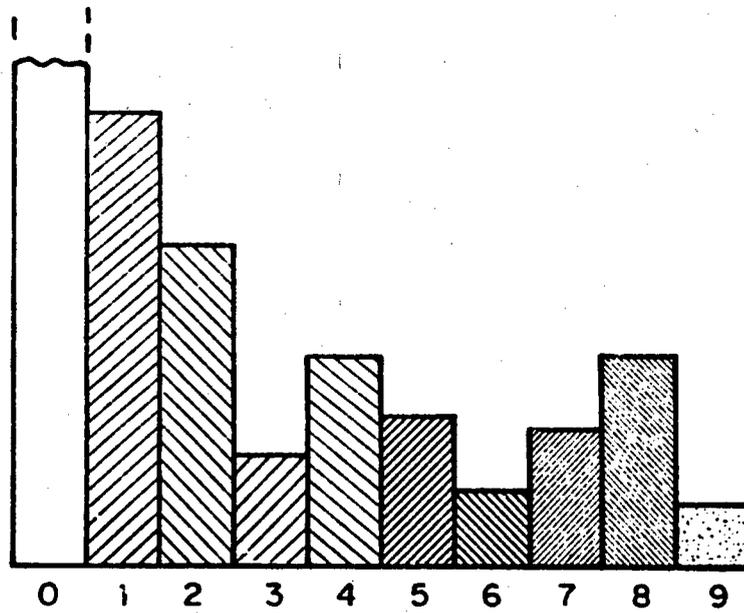


Figure 9. Frequency of evaluation numbers used by different evaluators (sample).

(A) ORIGINAL EVALUATIONS



(B) SMOOTHED EVALUATIONS, ADJUSTED TO AN IDEAL CURVE

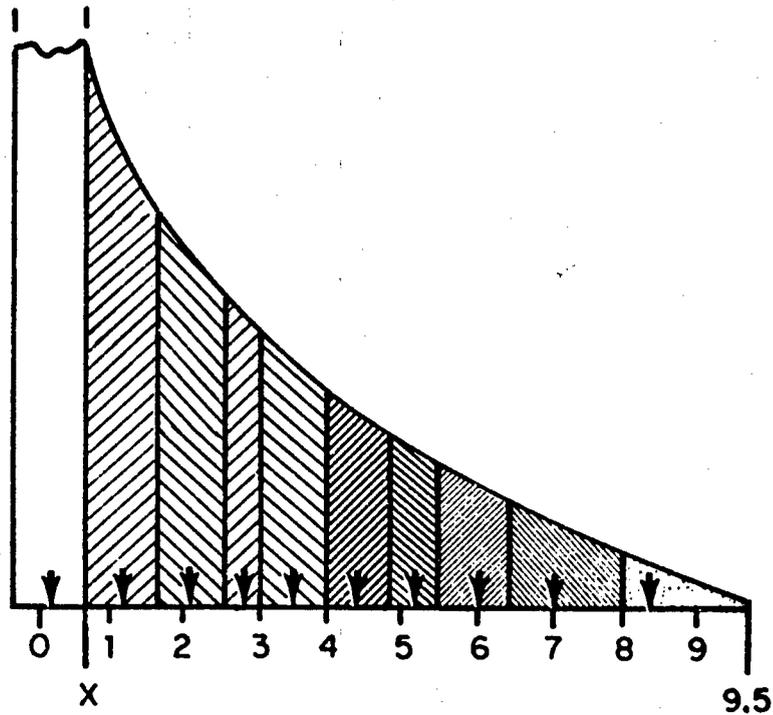


Figure 10. First smoothing method applied to the vectors of one evaluator.

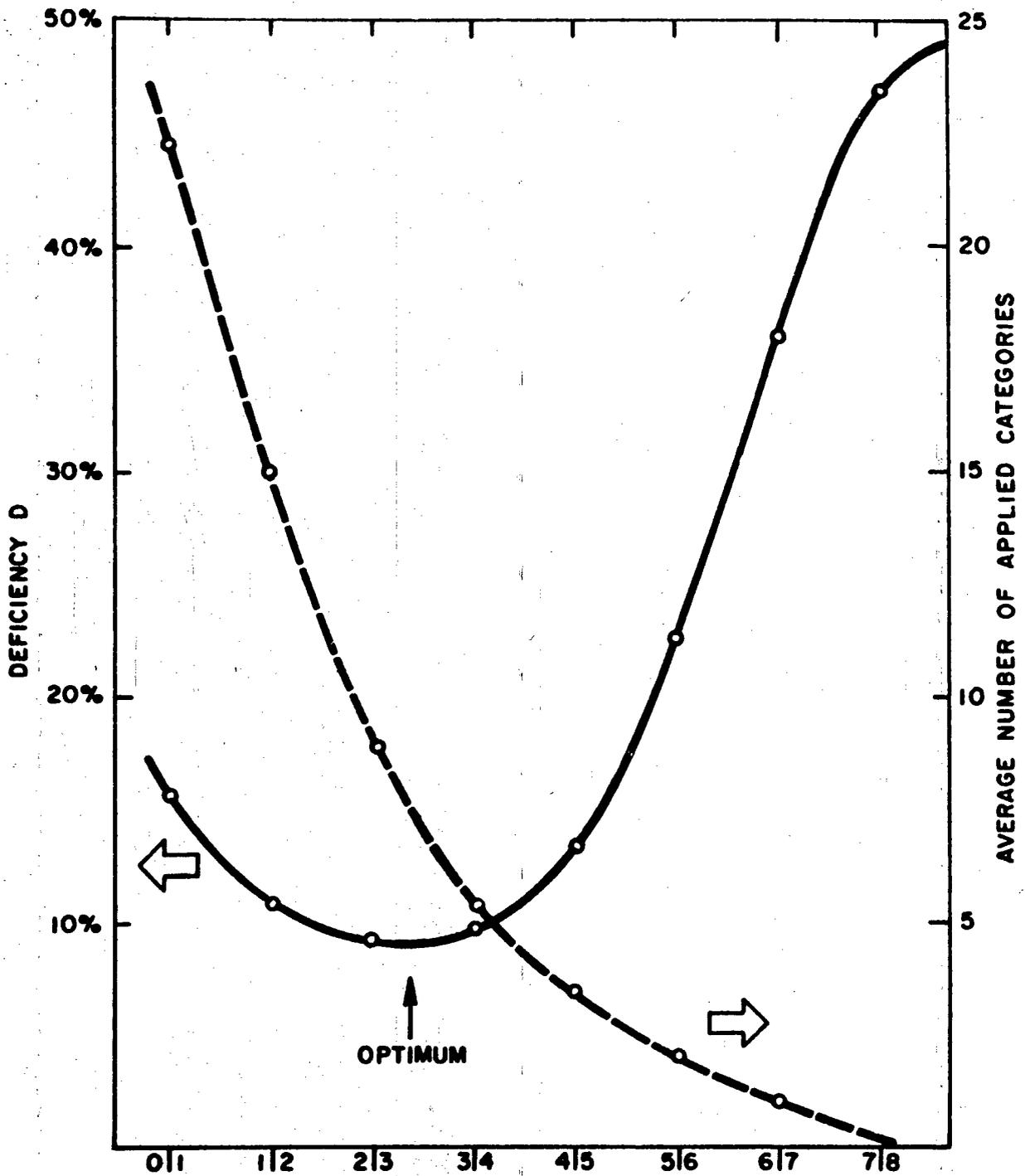


Figure 11. Effect of reducing the 10-valued to a 2-valued scale upon the deficiency (The circles on the solid line identify the divisions. The corresponding circles on the dashed line indicate the average numbers of the remaining categories having the value 1.)

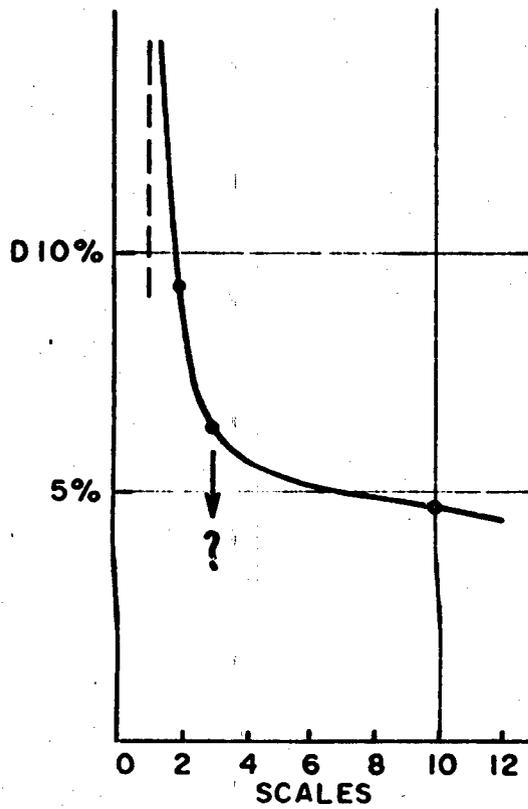


Figure 12. Effect of number of evaluation grades upon D.

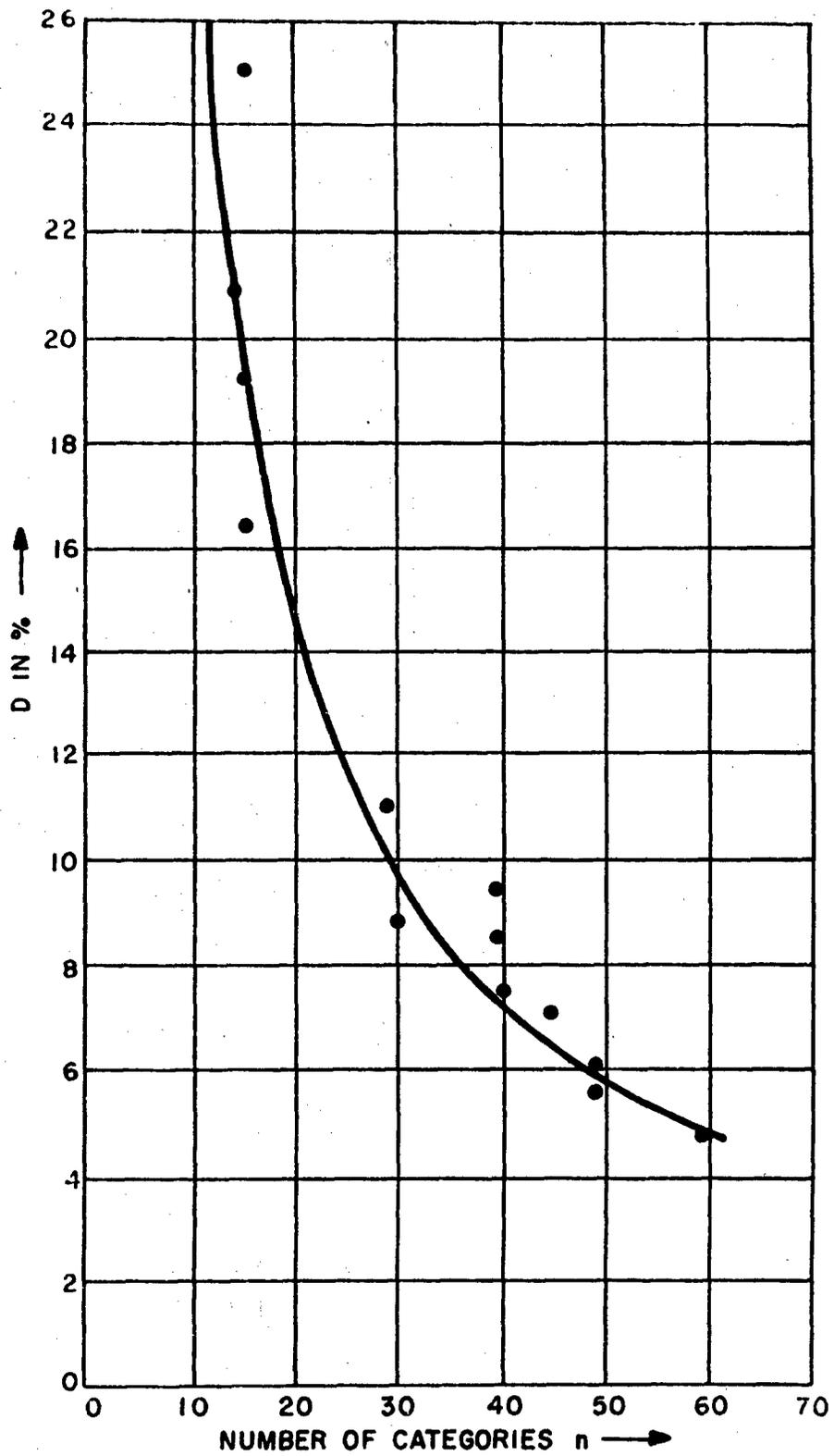


Figure 13. D vs number of applied categories.

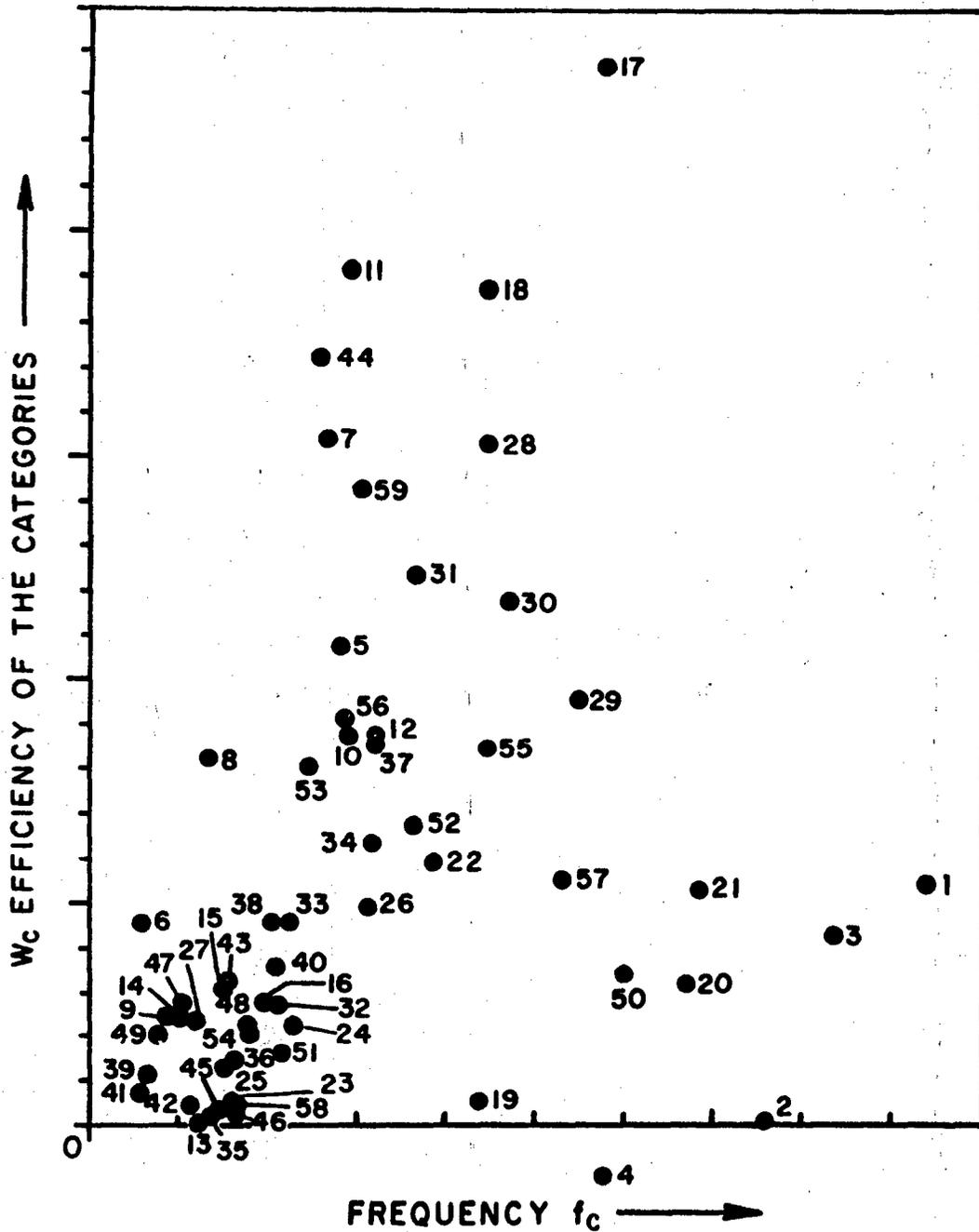


Figure 11. The effectiveness of the individual categories vs their frequency within the document vectors.

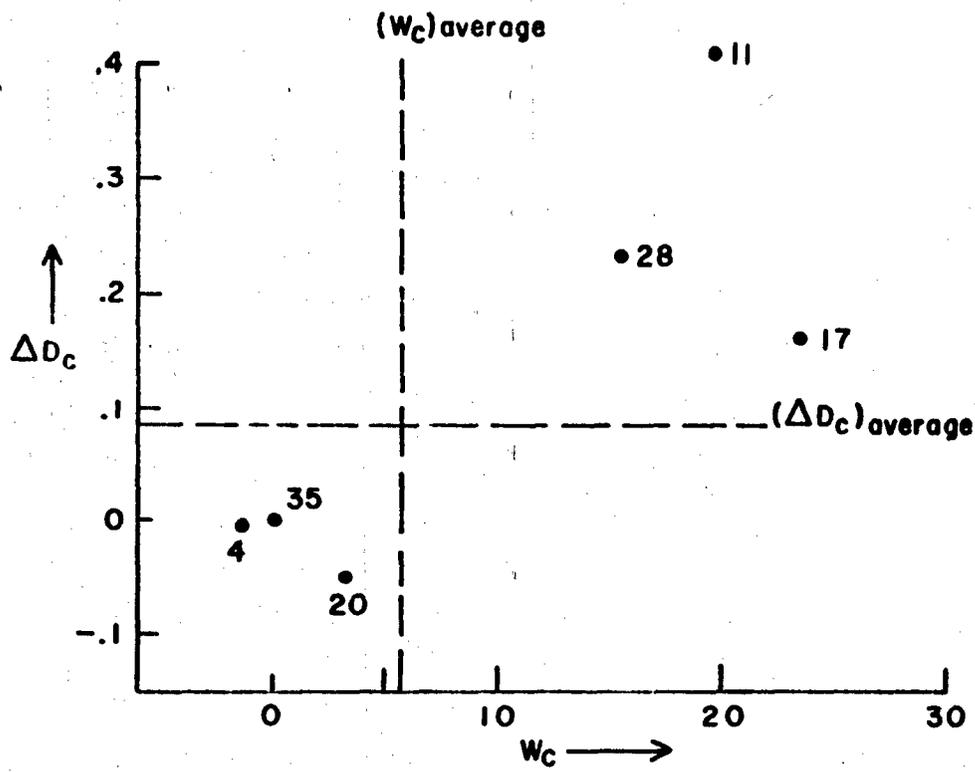


Figure 15. ΔD_c (change of D by dropping category c) versus W_c (efficiency of category c).

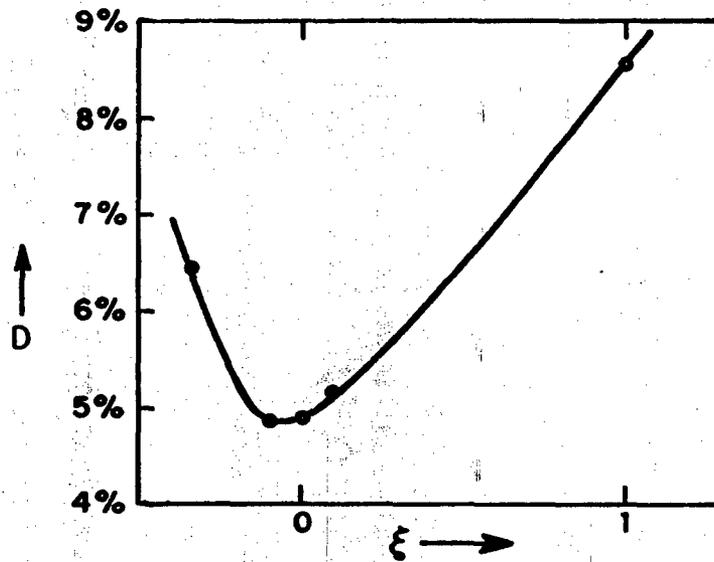


Figure 16. Changes of D when retrieval formula was modified by weight factors $W_c \xi$.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Harry Diamond Laboratories Washington, D. C. 20338		2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
		2b. GROUP ---	
3. REPORT TITLE AUTOMATION OF THE ABC SYSTEM			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)			
5. AUTHOR(S) (First name, middle initial, last name) Berthold Altmann Walter A. Riessler			
6. REPORT DATE August 1968	7a. TOTAL NO. OF PAGES 103	7b. NO. OF REFS 60	
8a. CONTRACT OR GRANT NO.		8b. ORIGINATOR'S REPORT NUMBER(S) TR-1392	
8c. PROJECT NO. DA-2P020401A728		8d. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
8e. AMCMS Code: 5910.21.63071			
8f. HDL Proj. No. 01200			
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.			
11. SUPPLEMENTARY NOTES ---		12. SPONSORING MILITARY ACTIVITY ARO	
13. ABSTRACT To advance the ABC system toward the automation of its retrieval and analytical input operations, linguistic problems were studied, and a prototype computerized retrieval test was conducted. A vector-type organization was imposed on the test collection. An appropriate measuring tool was constructed and used (a) to evaluate a variety of system parameters (ca 50 test runs were required) and (b) to rate different systems that evolve from the basic ABC model. The process of computerizing the standardization of ABC descriptors as well as the production of a comprehensive thesaurus (presenting terminology with associations and functions) are described and so are the methods prepared for progressive automation of the analytical effort in future test models.			

DD FORM 1473
1 NOV 65

REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED

Security Classification

14

KEY WORDS

LINK A

LINK B

LINK C

ROLE

WT

ROLE

WT

ROLE

WT

ABC storage and retrieval method

Storage and retrieval systems

- 1. Automation
- 2. Tests and testing
- 3. Evaluation models

Coordinate index method

Vector - type storage and retrieval systems

Operations analysis

Modeling

Documentation