Patterns of Thinking in Searching Patent Applications by Manual and Machine-Assisted Methods*

JUNE ROBERTS CORNOG
Institute for Applied Technology, National Bureau of Standards, Washington, D. C.

and PAMELA POWELL ELLIS
Research and Development, U. S. Patent Office, Washington, D. C.

Received August 18, 1965

INTRODUCTION

The United States Patent Office has found itself in an increasingly difficult position in the last decade. The rapidly expanding world technology has produced inventive ideas at an uncomfortable rate. The number of applications for patents, as well as patents granted, has risen until at present about 3,000,000 patents have been issued and that number increases by 60,000 every year.

Classification of patents into categories so they may be filed together and located by subject matter is more difficult than meets the eye at first; inventions are inventions because they are new and the fitting of new ideas or concepts into old classifications often causes stretching of the categories to include patents that do not seem to fit any too well.

When an inventor or designer has what he believes is a new idea, he and his lawyer draft an application for a patent on the invention. The application includes a specification which usually consists of a general background and description of the inventive idea and appropriate examples of its use, a statement of claims over areas covered by the idea and drawings, if necessary. The application for a patent is then submitted to the U. S. Patent Office where an examiner, a person presumed to be "skilled in the art," reads and evaluates it. The examiner then searches the prior art (previously issued U. S. and foreign patents, scientific literature, etc.) to determine whether the claimed invention is really novel and unobvious.

The examiner and the patent system are governed by a code of federal laws. The reason for the examiner's search is that patent statutes place limitations on what can be patented; the idea must be truly inventive. Insofar as novelty is concerned, the search presents a "yes–no" question, although the decision still has to be made by the examiner. It is in determining whether a thing is "obvious to one skilled in the art" that complex patterns of thinking are required.

To conclude in his own mind whether or not an idea is obvious, the examiner must use logical reasoning. He must not only be convinced himself that the idea is or is not inventive, but he must prepare the statement of his decision so that he can defend it to the applicant and, occasionally, to the Patent Office Board of Appeals. He depends on the references he finds in his search to establish the premises on which he will base his reasoning. The decision as to the obviousness of the inventive idea is considered by examiners generally to be the most delicate and professional task they perform. It requires the most sensitive handling, the finest analytical judgement, and the best possible supportive documents from the prior art.

Until approximately eight years ago, all applications for patents were accorded only a manual search of the prior art to determine whether the proposed idea was truly inventive. In making a search of the prior art the examiner first decided in what classifications the general idea was likely to be filed and he then found those "shoes" or file drawers and scanned the contents. Various short-cut helps were devised through the years; previous searchers would have put penciled notes on the old patents; "briefs" or abstracts were written for patents in some files, etc. But in the main the examiner was, and is, expected to scan the prior art in a category, determine rapidly any relevance to the principle in the application, and ultimately decide upon the inventiveness of the applicant's idea. If the prior art is extensive, or if the new idea is classifiable in any of several different categories, then the examiner will be ahead if he is familiar enough with the art to *know* where the most relevant art is and what the principal disclosures in the area have been. The brain of the examiner has functioned as an information storage system for the 175 years of Patent Office existence but the amount to be stored now doubles every few years.

The Patent Office has been attempting to meet the technical information explosion by instituting mechanical storage of the prior art. The enormous complexity of this problem will not be described here except to point out what may not be obvious to everyone: because of both human and machine incapacity to store information from whole disciplines at once, it has been necessary to set up numerous small systems, each limited to a particular segment of the art, with attendant danger of incompatibility among them. Several good small systems have been designed, however, and one of them is the subject of the analysis presented in this paper.

Designers of mechanical information storage systems should consider these two guiding philosophies: (1) the system should offer the user a "natural" method of querying the machine; and (2) the human nervous system should be used as a model insofar as possible since it

* Presented before the Division of Chemical Literature, Symposium on Work and Time Studies in Technical Information, 149th National Meeting of the American Chemical Society, Detroit, Mich., April 8, 1965.

appears to be the most efficient information retrieval system yet discovered. These guide lines are particularly appropriate in the design of retrieval systems for the Patent Office, owing to the peculiar nature of the work involved. It is regrettable that the true functioning of the nervous system thus far has pretty well remained a secret. One of the NATO divisions has, as the goal of its annual conference, the sharing of the latest nervous system physiological information and theories among the member nation scientists. The purpose of the clearinghouse conference is to seek ways of applying any promising discoveries to the design of mechanized systems of information storage and retrieval. The annual NATO conference is just one international effort to meet an information availability problem that has alarmed the world.

One complaint made by examiners about mechanized retrieval in general is that it forces the searcher into such a stylized way of thinking that he feels restrained by the system or at least unable to explore and develop new approaches to solving his problem. It is, at the least, different from the manual system he ordinarily uses to find information. The very presence of a mechanical contrivance between him and the thing he wishes to find seems in many instances to give him pause and a slight case of inertia about becoming involved in extensive or repeated inquiries. As a very simple illustration, a possibly similar reduction of creativeness often occurs when a person who can type relatively well must compose a paper calling for his best efforts. He finds his thoughts refuse to form when they must reach the paper through the medium of the typewriter. He is frustrated by the mechanics of operating the machine and turns to the older, slower pencil because it interposes fewer obstacles between his thoughts and their formalization.

Again, is it not possible that the "natural" way of searching for information or of putting one's thoughts on paper with a minimum of intermediary interference is easiest or most congenial simply because it was learned first? Will children who learned to use a typewriter as they learned their alphabets think just as constructively through the medium of the machine? Will future generations find a computer system more "natural" than the drawer of library index cards?

In the organometallics art of the U. S. Patent Office, the information from a somewhat limited group of U. S. patents is stored in a punched card system which is searched by sorting equipment. Because only U. S. patent information is filed in the mechanized deck, examiners regularly search the prior U. S. art by machine but use the manual system for foreign references and other pertinent material. The regular practice of both skills renders these examiners almost unique in the examining corps.

In general, the organometallics mechanized file can be described as a classification of the art by both conceptual ideas and fragments or details, which may or may not be applicable to more than one concept. When an examiner wishes to address an inquiry to the system, he makes out multiple levels—broad or generic concepts, intermediate and fragmentary descriptors—and the machine will return to him lists of the numbers of patents which contain information relevant to the application being examined. The exact definition of each of these types of questions will be clarified by example when we use a case as illustration.

PROBLEM

Let us turn to the problem which is the central issue of this paper. The Patent Office asked for an unbiased, objective study of the differences in patterns of thinking associated with manual and mechanized searches during an examination of the prior art. The complete processing or disposal of an application was not included in the assignment.

PROCEDURE

To essay a study of mental processes, it was necessary to use a highly subjective, introspective approach because the researcher is limited in this type of inquiry to two avenues of analysis: first, what the examiner himself can tell you about what he is thinking, and second, a comparison of his verbal material with the actions he takes in going through the search.

In all, 12 protocols were recorded; six cases were examined both manually and by machine. Three examiners participated, two in the organometallics art and one in the electronics transistor art. Each of the three examiners made a total of four searches, two manual and two machine.

The searches were performed by Mr. Tobias Levow, Primary Examiner in Group 110, Organo-Metallics, U. S. Patent Office, and Mrs. Ernestine Bartlett, examiner in the same art. Although the work of the electronics examiner, Mr. Donald Forrer, will not be reported more than incidentally in this paper, the findings with respect to his searches correspond closely to those of Mr. Levow and Mrs. Bartlett.

Only one case, searched both manually and by machine by Mr. Levow, will be presented as an illustration of the principal findings, owing to the detailed nature of the analysis given here. The procedure and contrasts are typical of those found in the work of the other examiners in this extremely small-sample, highly subjective inquiry. It should also be added that, in the interests of Patent Office security, applications were used which had actually been issued as patents at least one year before the beginning of the study. The patents actually used serve only to illustrate the search procedures and are not of importance in themselves for the import of the study.

The original forms of the applications for the base patents used in the study were brought out of the files and any notes about amendments, etc., were deleted in the photographic reproduction. All manual and machine searches using the same application were made at least two months apart to allow time for the examiner to forget exactly what he had found in the previous search. The area of search was limited, furthermore, to U. S. patents, since these were the only ones that had been stored in the mechanized file.

THE MECHANIZED SEARCH

The investigation began with the mechanized search of an application for a patent concerning complex-forming polyaminopolycarboxylic acids. The planning and execution of the machine-assisted search will be traced through to completion first, and then followed by a similar analysis of the manual search for the same case.

The application was chosen for use in the specimen searches only because it was relatively uncomplicated; it had no other especially interesting or convenient characteristics. It concerned compounds which differed from other known compounds *only* by the presence of a different hydrocarbon group from that present in the old compounds. The inventor stated that this small change gave unexpected results, ones not "obvious to one skilled in the art." In both the manual and machine searches, the examiner began by reading the specification, the general description of and background for the claimed invention. He then turned to and read carefully the claims section where the inventor sets forth the exact statement of the areas he claims are covered by his invention. The examiner's own statement of the principle around which the application centered is enlightening, but the reader should keep in mind that it is a verbatim quotation of a man talking essentially to himself and that spoken phraseology is never as precisely stated as written:

"This case relates to allegedly novel polyaminopolycarboxylic acids and their metal chelates. It continues particularly with alicyclic triamino pentaacetic acids, which are stated to exhibit sequestering activity and are useful as chelating agents. It states further that the invention also pertains to the new starting materials used in the production of the subject compounds, although what these 'new starting materials' are I can't tell at this point. The case acknowledges the general class of alkaline, diamine, acetic acid chelates or chelating agents, ethylenediaminetetraacetic acid, diethylenetriaminopentaacetic acid, and cyclohexyldiaminotetraacetic acid. These are the known, outstanding members of the group. He talks about areas of use of these compounds. He identifies these as N,N′,N′,N″,N″*-pentacarboxymethyl-N-B-(aminoethyl)-1,2-diaminocycloalkanes. Basically this is a compound of the type very much like the diethylenetriaminopentaacetic acid that was mentioned as prior art, except that instead of one of the ethylene groups there is a cycloalkyl group, which is further stated as being 1,2-disubstituted. Otherwise the compound is identical with the acknowledged compound. This may pose a serious problem not in posing a search question but in the search procedure we have set up, being able to differentiate effectively between this compound and the acknowledged prior art. The only thing we had in the referenced document is the compound named as prior art which the search procedure would differentiate readily. The problems will arise in distinguishing when the reference shows a number of different materials because the search procedure does 'compositing' and can give us a lot of references that really don't bear much relationship to the problem." (*an exact quotation.)

As he read the specification, the first characteristic of machine search appeared: he isolated the *concept* central to the claimed invention, then followed with the small but specific details associated with that concept in this case, and finally structured these in a way he knew the machine could use in look-up. He prepared to search the concept, evidently hoping to find a reference as close to being "pat" as possible with his first set of queries. Figures 1 and 2 show the method of stating the questions for the machine search.

The entire search was planned before the examiner looked at a single reference. He decided on a broad concept question, an intermediate level question, and a specific question for each of the two main channels of inquiry, cyclopentyl and cyclohexyl, and he tried to limit the number of parameters on each of the six question sheets to not over 12 items in order not to require the machine tabulation people at the data processing branch to have to plug extra boards for the sorting equipment. At the same time he tried to cover as many aspects as would be likely to be fruitful so he would not have to address a second set of questions to the machine room.

He drafted the machine queries, using the code sheet prepared by the system designers, as shown in Figures 1 and 2.

First, the fragment connections (Figure 1)
  Under Alkylene
    COOR
    Amine
  Under Cycloalk
    Amine
Second, the fragment descriptions (Figure 2)
  Under Cycloalkyl
    6M
  Under Amine
    2+ (2 or more)
    Tertiary
Under 2+ identical fragment connections
    Alkylene
    Cycloalk
    COOR
    Amine

The intermediate level question for the cyclohexyl concept specified a metal chelate, while the third and most specific question asked for an iron chelate. The questions for cyclopentyl were equivalent to those for cyclohexyl.

The data processing branch supplied the applicable references as lists of patent numbers. The examiner began with the shortest lists, those which came in answer to the two most specific questions. He compared the patent numbers "dropped out" in answer to questions 5 and 6, the iron chelate for cyclohexyl and cyclopentyl, and eliminated all duplicates. He then went to the "shoes" or file drawers of issued patents and drew out those listed by the machine tabulation. These he took back to his desk and examined in detail.

Without further description, a synopsis of the number of references supplied by the machine is

	cyclohexyl	cyclopentyl
Broad question	47	16
Intermediate question	25	7
Specific question	13	3

The lists of patents cited by the machine telescoped within each major channel of inquiry; each next broader level contained all the references mentioned in the more specific levels below it. It was also interesting to note that the patent issued on this application was one of the references on each of the six answer sheets. (It was gratifying to all concerned to have this testimony to the accuracy of the coding and the relevance of the examiner's question formulation.)

Of the 25 different patents listed by the machine in answer to the two more specific questions on the cyclohexyl group, the examiner read or skimmed 16 and laid

ORGANOMETALLIC CODE SHEET

FRAGMENT — CONNECTIONS

ALKYL

	0	Halo
	1	Aryl
	2	Cycloalk
	3	COOR
	4	S-COOR
	5	Metal
	6	S-Hetero
	7	O-Hetero
	8	N-Hetero
	9	N,C,S
	11	N,C,O
1	12	S,O
2	0	C,N
	1	Carbonyl
	2	N,O
	3	Amine
	4	O,(S)
	5	Phosphorus
	6	OH,(SH)
	7	Miscell.

ALKYLENE

	8	Halo
	9	Aryl
	11	Cycloalk
2	12	COOR
3	0	S-COOR
	1	Metal
	2	S-Hetero
	3	O-Hetero
	4	N-Hetero
	5	N,C,S
	6	N,C,O
	7	S,O
	8	C,N
	9	Carbonyl
	11	N,O
3	12	Amine
4	0	O,(S)
	1	Phosphorus
	2	OH,(SH)
	3	Miscell.

ALKINYL

	4	Halo
	5	Aryl
	6	Cycloalk
	7	COOR
	8	S-COOR
	9	Metal
	11	S-Hetero
4	12	O-Hetero
5	0	N-Hetero
	1	N,C,S
	2	N,C,O
	3	S,O
	4	C,N
	5	Carbonyl
	6	N,O
	7	Amine
	8	O,S
	9	Phosphorus
	11	OH,(SH)
5	12	Miscell.

ALKENYL

	0	Halo
	1	Aryl
	2	Cycloalk
	3	COOR
	4	S-COOR
	5	Metal
	6	S-Hetero
	7	O-Hetero
	8	N-Hetero
	9	N,C,S
	11	N,C,O
6	12	S,O
7	0	C,N
	1	Carbonyl
	2	N,O
	3	Amine
	4	O,(S)
	5	Phosphorus
	6	OH,(SH)
	7	Miscell.

HALOGEN

	8	Aryl
	9	Cycloalk
	11	COOR
7	12	S-COOR
8	0	Metal
	1	S-Hetero
	2	O-Hetero
	3	N-Hetero
	4	N,C,S
	5	N,C,O
	6	S,O
	7	C,N
	8	Carbonyl
	9	N,O
	11	Amine
8	12	O,(S)
9	0	Phosphorus
	1	OH,(SH)
	2	Miscell.

ARYL

	3	Aryl
	4	Cycloalk
	5	COOR
	6	S-COOR
	7	Metal
	8	S-Hetero
	9	O-Hetero
	11	N-Hetero
9	12	N,C,S
10	0	N,C,O
	1	S,O
	2	C,N
	3	Carbonyl
	4	N,O
	5	Amine
	6	O,(S)
	7	Phosphorus
	8	OH,(SH)
	9	Miscell.

MISCELLANEOUS

24	11	Miscell.

OH,(SH)

24	9	Miscell.

PHOSPHORUS

	6	Phosphorus
	7	OH,(SH)
24	8	Miscell.

O,(S)

	2	O,(S)
	3	Phosphorus
	4	OH,(SH)
24	5	Miscell.

AMINE

	9	Amine
	11	O,(S)
23	12	Phosphorus
24	0	OH,(SH)
	1	Miscell.

N,O

	3	N,O
	4	Amine
	5	O,(S)
	6	Phosphorus
	7	OH,(SH)
23	8	Miscell.

CYCLOALK

	11	Cycloalk
10	12	COOR
11	0	S-COOR
	1	Metal
	2	S-Hetero
	3	O-Hetero
	4	N-Hetero
	5	N,C,S
	6	N,C,O
	7	S,O
	8	C,N
	9	Carbonyl
	11	N,O
11	12	Amine
12	0	O,(S)
	1	Phosphorus
	2	OH,(SH)
	3	Miscell.

COOR

	4	S-COOR
	5	Metal
	6	S-Hetero
	7	O-Hetero
	8	N-Hetero
	9	N,C,S
	11	N,C,O
12	12	S,O
13	0	C,N
	1	Carbonyl
	2	N,O
	3	Amine
	4	O,(S)
	5	Phosphorus
	6	OH,(SH)
	7	Miscell.

S-COOR

	8	S-COOR
	9	Metal
	11	S-Hetero
13	12	O-Hetero
14	0	N-Hetero
	1	N,C,S
	2	N,C,O
	3	S,O
	4	C,N
	5	Carbonyl
	6	N,O
	7	Amine
	8	O,(S)
	9	Phosphorus
	11	OH,(SH)
14	12	Miscell.

METAL

	0	Metal
	1	S-Hetero
	2	O-Hetero
	3	N-Hetero
	4	N,C,S
	5	N,C,O
	6	S,O
	7	C,N
	8	Carbonyl
	9	N,O
	11	Amine
15	12	O,(S)
16	0	Phosphorus
	1	OH,(SH)
	2	Miscell.

S-HETERO

	3	S-Hetero
	4	O-Hetero
	5	N-Hetero
	6	N,C,S
	7	N,C,O
	8	S,O
	9	C,N
	11	Carbonyl
16	12	N,O
17	0	Amine
	1	O,(S)
	2	Phosphorus
	3	OH,(SH)
	4	Miscell.

O-HETERO

	5	O-Hetero
	6	N-Hetero
	7	N,C,S
	8	N,C,O
	9	S,O
	11	C,N
17	12	Carbonyl
18	0	N,O
	1	Amine
	2	O,(S)
	3	Phosphorus
	4	OH,(SH)
	5	Miscell.

CARBONYL

	8	Carbonyl
	9	N,O
	11	Amine
22	12	O,(S)
23	0	Phosphorus
	1	OH,(SH)
	2	Miscell.

C,N

	0	C,N
	1	Carbonyl
	2	N,O
	3	Amine
	4	O,(S)
	5	Phosphorus
	6	OH,(SH)
	7	Miscell.

S,O

	3	S,O
	4	C,N
	5	Carbonyl
	6	N,O
	7	Amine
	8	O,(S)
	9	Phosphorus
	11	OH,(SH)
21	12	Miscell.

N,C,O

	5	N,C,O
	6	S,O
	7	C,N
	8	Carbonyl
	9	N,O
	11	Amine
20	12	O,(S)
21	0	Phosphorus
	1	OH,(SH)
	2	Miscell.

N,C,S

	6	N,C,S
	7	N,C,O
	8	S,O
	9	C,N
	11	Carbonyl
19	12	N,O
20	0	Amine
	1	O,(S)
	2	Phosphorus
	3	OH,(SH)
	4	Miscell.

N-HETERO

	6	N-Hetero
	7	N,C,S
	8	N,C,O
	9	S,O
	11	C,N
18	12	Carbonyl
19	0	N,O
	1	Amine
	2	O,(S)
	3	Phosphorus
	4	OH,(SH)
	5	Miscell.

Figure 1. Code sheet as it was sent to machine tabulation, showing items marked in asking the generic question on fragment connections.

FRAGMENT DESCRIPTIONS

ALKYL
0 Alkyl
1 Lo (1-7C)
2 1C
3 2C
4 3C
5 4C
6 5-7C
7 Hi (8+C)
8 8-12C
9 13+C

ALKYLENE
11 Alkylene
25/26 12 Lo (1-7C)
0 1C
1 2C
2 3C
3 4C
4 5-7C
5 Hi (8+C)
6 8-12C
7 13+C

ALKINYL
8 Alkinyl
9 Lo (1-7C)
11 2C
26/27 12 Hi (8+C)
0 Mono =
1 Poly =

ALKENYL
2 Alkenyl
3 Lo (1-7C)
4 =CH$_2$
5 2C
6 3C
7 4C
8 5-7C
9 Hi (8+C)
11 8-12
27/28 12 13+C
0 Mono =
1 Poly =

HALOGEN
2 Halo
3 F
4 Cl
5 Br
6 I
7 1 Halo
8 2 Halo
9 3 Halo
11 4 Halo
28/29 12 5+Halo

ARYL
0 Aryl
1 2+
2 1 Sub
3 2 Sub
4 D (O)
5 E (M)
6 F (P)
7 3 Sub
8 G (O,M)
9 H (O,M,P)
11 J (M)
29/30 12 4 Sub
0 5 Sub
1 6 Sub
2 Ind
3 Spiro
4 Fused

CYCLOALKYL
5 Cycloalk
6 2
7 3+
8 Saturated
9 Unsat.
11 3,4M
30/31 12 5M
0 diene
1 6M
2 7+M
3 Bicyc (1C)
4 Bicyc (2C)
5 Ind
6 Spiro
7 Fused

COOR
8 COOR
9 2+
11 R is H
31/32 12 R is not H
0 COHalo
1 OCOO
2 Anhydride
3 Miscell.

S-COOR
4 S-COOR
5 2+
6 CSSH
7 CSSR
8 CSOH
9 CSOR
11 COSH
32/33 12 COSR
0 CSHalo
1 XCXX
2 Anhydride
3 Miscell.

S-HETERO
4 S-Hetero
5 3,4M
6 5M
7 6M
8 7+M
9 O-Cont
11 N-Cont
33/34 12 Other Cont
0 1S
1 2+S
2 Ind
3 Spiro
4 Fused

N,O
5 NO
6 NO$_2$
7 NOH
8 Miscell.

O-HETERO
9 O-Hetero
11 3,4M
34/35 12 5M
0 6M
1 7+M
2 N-Cont
3 S-Cont
4 Other Cont
5 1-O
6 2+O
7 Ind
8 Spiro
9 Fused

N-HETERO
11 N-Hetero
35/36 12 3,4M
0 5M
1 6M
2 7+M
3 O-Cont
4 S-Cont
5 Other Cont
6 1N
7 2N
8 3+N
9 Salt
11 Ind
36/37 12 Spiro
0 Fused

CARBONYL
1 C=O
2 CHO
3 C=S
4 CHS
5 =O (ring)
6 =S (ring)
7 Metal Carb
8 1 CX
9 2 CX
11 3 CX
37/38 12 4 CX
0 5 CX
1 6+CX

N,C,O
2 N-C(=O)-
3 N-C(=O)-O-
4 N-C(=O)-N
5 CNO
6 NCO
7 Miscell.

S,O
8 SO
9 SO$_2$
11 SO$_3$H
38/39 12 SO$_4$
0 Miscell.

C,N
1 CN
2 NC
3 N-C(=N)-
4 N-C(=N)-N
5 Miscell.

N,C,S
6 N-C(=S)-
7 N-C(=S)-S-
8 N-C(=S)-N
9 SCN
11 CNS
39/40 12 Miscell.

OH,(SH)
0 OH
1 SH
2 1 XH
3 2 XH
4 3 XH
5 4 XH
6 5+XH

AMINE
7 2+
8 Primary
9 Secondary
11 Tertiary
40/41 12 Quaternary
0 Diazo
1 Azide
2 Nitride
3 =NH
4 =NR
5 Hydrazine
6 Salt
7 Miscell.

O,(S)
8 -O-
9 -O-O-
11 -S-
41/42 12 -S-S-
0 -(S)$_{3+}$-
1 2 X
2 3 X
3 4+X

PHOSPHORUS
4 X=PX$_3$
5 1+S
6 4S
7 =S,2S,O
8 =S,S,2O
9 =S,3O,
11 =O,3S
42/43 12 =O,3O
0 =O,2S,O
1 =O,S,2O
2 P-Other
3 P-X-P
4 Poly-P
5 Cyclic-P
6 P+3
7 P+5

MISCELLANEOUS
8 Ferrocene
9
11
12 Miscell.

METALS
0 I-A
1 Li
2 Na
3 K
4 Rb
5 Cs
6 II-A
7 Be
8 Mg
9 Ca
11 Sr
44/45 12 Ba
0 III-A
1 B
2 Al
3 Ga
4 In
5 Tl
6 IV-A
7 Si
8 Ge
9 Sn
11 Pb
45/46 12 V-A
0 As
1 Sb
2 Bi
3 VI-A
4 Se
5 Te
6 Po
7 I-B
8 Cu
9 Ag
11 Au
46/47 12 II-B
0 Zn
1 Cd
2 Hg
3 III-B
4 Sc
5 Y
6 La
7 IV-B
8 Ti
9 Zr
11 Hf
47/48 12 V-B
0 V
1 Nb
2 Ta
3 VI-B
4 Cr
5 Mo
6 W
7 VII-B
8 Mn
9 Tc
11 Re
48/49 12 VIII-B
0 Fe
1 Ru
2 Os
3 Co
4 Rh
5 Ir
6 Ni
7 Pd
8 Pt
9 Lanthanide
11 Ce
49/50 12 Other
0 Actinide
1 Th
2 U
3 Other
4 Hydride
5 Bis Alkyl
6 Tris Alkyl
7 Tetra Alk
8 Unkn. Conn
9 Cyc Metal
11 Chelate
50/51 12
0 Mono Met
1 Poly Met
2 2 M
3 3 M
4 4 M
5 5+M
6 Mixed Met
7 2 M
8 3 M
9 4+M
11
12

# of Metal Connections
0 1 Conn.
1 2 Conn.
2 3 Conn.
3 4 Conn.
4 5 Conn.
5 6+Conn.

# of Identical Metal Conn.
6 1 I.C.
7 2 I.C.
8 3 I.C.
9 4 I.C.
11 5 I.C.
52/53 12 6+I.C.

2+ Identical Frag. Connections
0 Alkyl
1 Alkylene
2 Alkinyl
3 Alkenyl
4 Halo
5 Aryl
6 Cycloalk
7 COOR
8 S-COOR
9 Metal
11 S-Hetero
53/54 12 O-Hetero
0 N-Hetero
1 N,C,S
2 N,C,O
3 S,O
4 C,N
5 Carbonyl
6 N,O
7 Amine
8 O,(S)
9 Phosphorus
11 OH,(SH)
12 Miscell.

MISCELLANEOUS DICTIONARIES

ANION		P-OTHER		Misc N,C,S,O		MISCELL.	
55	56	57	58	59	60	61	62
0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9
11	11						
12	12						

RING DESCRIPTORS

Ring System (General)
0 2+R.S.
1 2 Rings
2 3 Rings
3 4 Rings
4 5 Rings
5 6 Rings
6 1 N-Ring
7 2 N-Rings
8 3 N-Rings
9 1 S-Ring
11 2 S-Rings
63/64 12 1 O-Ring
0 2 O-Rings
1 1 Mis-Het R
2 2 Mis-Het R
Fused Face
3 M
4 N
5 P
6 Q
7 R
8 S
9 T
11 U
12 V

RING POSITION #
0 #1
1 #2
2 #3
3 #4
4 #5
5 #6
6 #7
7 #8
8 #9
9 #10
11 #11
65/66 12 #12
0 #13
1 #14
2 #15
3 #16
4 #17
5 #18
6 #19
7 #20-24
8 #25-29
9 #30
11 Mono R
12 Poly R

RING INDEX #

67	68	69	70
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9

U.S. DEPT OF COMMERCE
Patent Office

R&D J.Frome & P.O'Day
December 1, 1961

DOCUMENT IDENTIFICATION
71 72 73 74 75 76 77

Analyst__________ Checked By__________

Figure 2. Code sheet as it was sent to machine tabulation, showing items marked in asking the generic question on fragment descriptions.

aside 7 as being useful in meeting the claims in the application. These 7 he informally classified as

Good combining art or reference	2
Possibly useful	2
Teaches equivalence	1
Pertinent	1
Most pertinent	1

In all, one-half of the machine-recommended references in answer to the most specific question about cyclohexyl had shown some usable degree of pertinence in meeting the claims of the application. Of these, one was accepted as being an almost anticipatory reference, with one other "combining" document added to ensure coverage of all aspects.

Now let us look at two charts of examiner thought. Figure 3 is a schematic presentation of the plan and execution of the mechanized search. It indicates the final incomplete state of five of the channels the examiner had originally thought he might follow to reach his goal of a reasonably pertinent reference, plus the termination, almost on exact target, of the one channel he found to be sufficient for his needs. Figure 4 indicates what happened to each of the sixteen patents in order for the examiner to end up with a final score of two accepted references. We have paid especial attention to the mental excursions he made and to any deviations from the search plan he laid down in starting. We also noted any flexibility he showed, any willingness to consider alternate combinations or methods, and any hesitation of any kind in his drive to find the most applicable references.

THE MANUAL SEARCH

Let us turn next to the manual search of the same application, performed by the same examiner two months later. He started, of course, by reading the specification and turning to the claims section as soon as he had the general idea of the invention in mind.

As he read the application, the examiner's thoughts concentrated, in the light of his intimate knowledge of the art, on where the most pertinent references were likely to be filed. When he had finished his analysis of the specification and claims, he designated four Patent Office subclasses he would search. He also indicated he would start with a certain unofficial subclass or digest which the examiners in the art had established themselves. His search plan was simple: he would start out to find the prior art that had been acknowledged in the application. He said he would then read those patents and decide what to do next. In his own words

> "We will limit the search to the heavy or polyvalent metals, the heavy metal of aluminum salts; ones which would be classified here in this search deck (U. S. Patents only)."

He was deliberately limiting the field of search to the U. S. patents which were covered by the mechanized file because he had been asked to do so. Continuing with his own statement of his search plan:

> "We know that we should be able to find to start with, the two compounds that he acknowledges as prior art. These are the diethylenetriaminopentaacetic acid and its chelates and the cyclohexyldiaminetetraacetic acid and its chelates; that one (of the

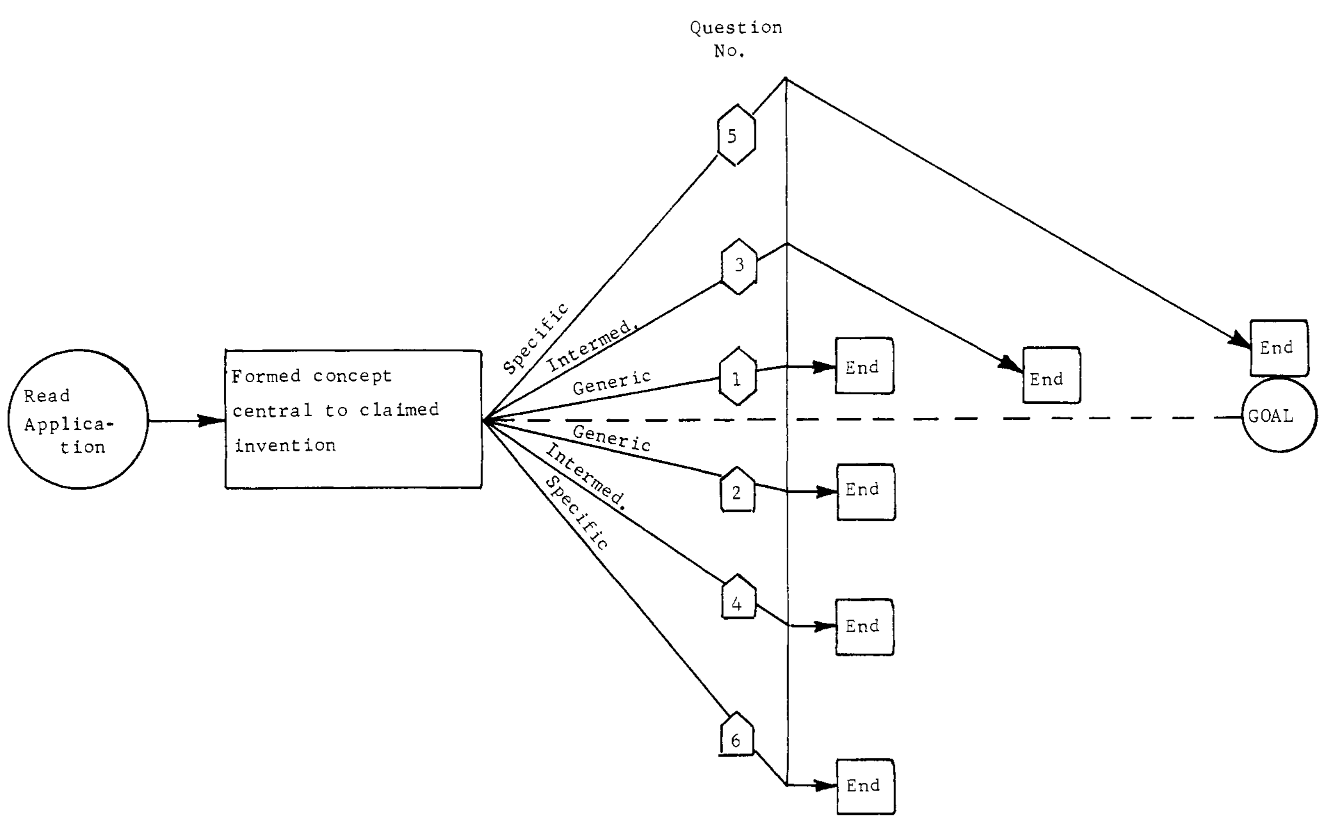

Figure 3. Schematic diagram of mechanized search plan and its execution.

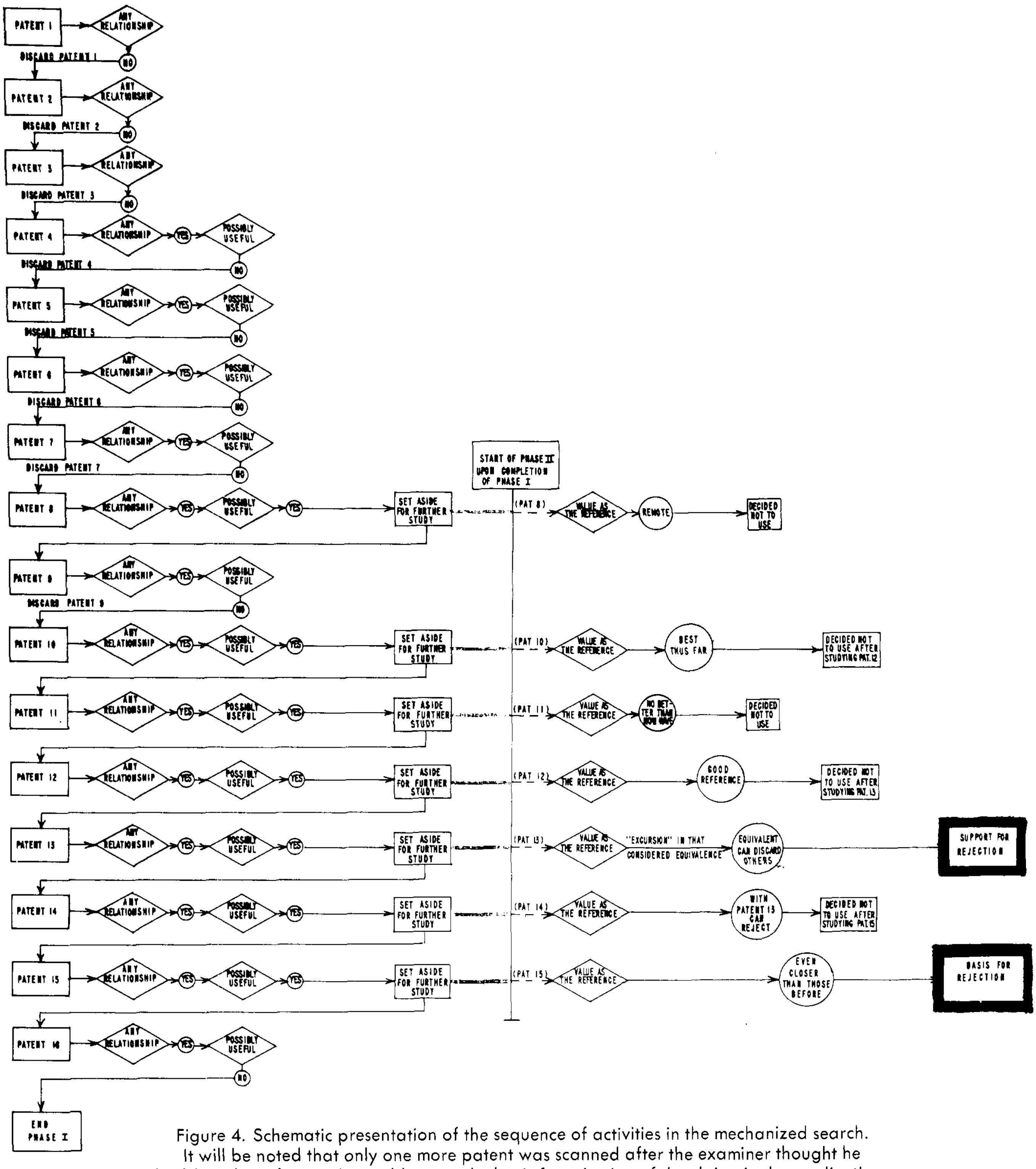

Figure 4. Schematic presentation of the sequence of activities in the mechanized search. It will be noted that only one more patent was scanned after the examiner thought he had found a reference he could use as the basis for rejection of the claims in the application.

above compounds) having the cyclohexyl function between two nitrogens is bonded to these acetic acid moieties. The former one has the same basic general structure as the applicant's except that it has only two ethylene groups where the applicant has a cyclohexyl group instead. So, we'll keep our eyes open specifically for this but we will be looking for the particular compound.

I'll now start the search and since we do have a digest on chelates, that would be the first place I would look. . . . I shall be constrained to start the search with the most recent material and work backward because I know this stuff to a certain extent and I know that in the older patents I will find the simpler chelates and the classical chelates."

In searching the one shoe containing the digest, he probably looked at 40 to 50 patents but was able to discard the majority of them by looking at the titles alone since applicants are required to indicate there the nature and general coverage of their inventions. He laid aside and actually read thoroughly 11 of the patents and roughly classified them as

Not pertinent	2
Some relevance	3
Similar to what we have	1
Good back-up	2
Close or very close	3

In the final analysis the examiner stated that he thought four references were ones he would use against the claims in the application.

Now let us look at a chart of his mental activities as the examiner went through the digest shoe. He did not find it necessary to go to any of the other three subclasses he had considered in planning his search since the art in the digest effectively covered the materials he would have found in the other subclasses. We are interested in his "excursions" or his readiness to explore variations from his originally stated requirements. Figure 5 is a diagrammatic presentation of the manual search plan and its execution. The flow chart indicated that after deciding to search for the acknowledged prior art he did so in terms of three specific compounds and succeeded in locating good basic reference documents showing them. At this point he determined that he needed a teaching (Patent Office jargon for "statement by the inventor") to show that the unsymmetrical arrangement of the alkylene and cycloalkylene groups in the same compound was obvious because their function in this type of compound was equivalent. He then looked for and located references which, combined, gave him grounds for tenable rejection of the claims.

Figure 6 presents the details of the manual search. Of the probably 40+ patents in the digest shoe, 11 were selected for thorough examination as to their relevance. The chart shows the discriminations and decisions made after the examinations were completed.

FINDINGS OF FACT AND INFERENCES

Inferences stated below should be regarded only as hypotheses upon which to base more quantified research. The sample used in the investigation was inadequate to provide a basis for generalization to a larger population without quantitative verification.

1. The mechanized search started by being completely planned while the manual had a definite but only interim objective. The mechanical search was fully planned before the first question was asked. The queries posed to the machine had logical relationship to each other and all had definite places in the search plan structure.

The manual search began in one rather tentative direction. The examiner said he would try to find the prior art mentioned in the application and after reading it he would decide where to look next. His plan was definite only in that he knew he was going to look for three specific compounds. He expanded and refined his search to concentrate on parts missing after finding pertinent material in the prior art. He then glanced through side issues and indicated he would look for equivalents if he could not find the presence of the two different groups in the same compound.

In the mechanized search the examiner cleaved to his original search plan steadfastly. He allowed his thoughts to stray to possible alternatives just once. He seemed to check each machine drop with the thought, "Is this the compound I asked for or isn't it?" If it was not, he then discarded it forthwith. He evidently did not expect the system to point out any equivalent concepts unless he specifically asked it to do so.

In the manual search, the examiner apparently said to himself, "Give me something to start with," and he treated each of the 11 selected references as a definite possibility. He set about analyzing each with a view to using any information that might be applicable. He continually reviewed his original plan and seemed to be trying to stretch or twist it to accommodate possible equivalent structures or procedures. Each patent read became a vigorous lesson in chemistry as he strove to match or substitute.

2. The basic conceptions of manual and machine search are somewhat different. In the manual system patent information is organized and classified on the basis of concepts. Certain principles of invention are common to the patents filed together. Actual terminology used in a patent, or specific details of processes listed in the claims, must be translated into one of the concepts included in the classi-

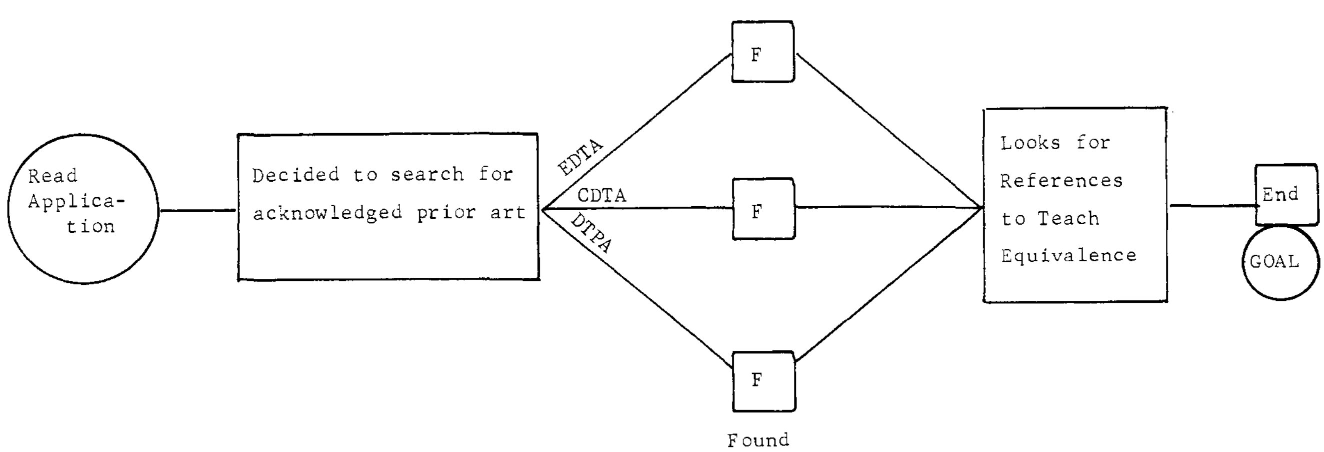

Figure 5. Schematic diagram of manual search plan and its execution: EDTA, ethylenediaminetetraacetic acid; CDTA, cyclohexyldiaminotetraacetic acid; DTPA, diethylenetriaminopentaacetic acid.

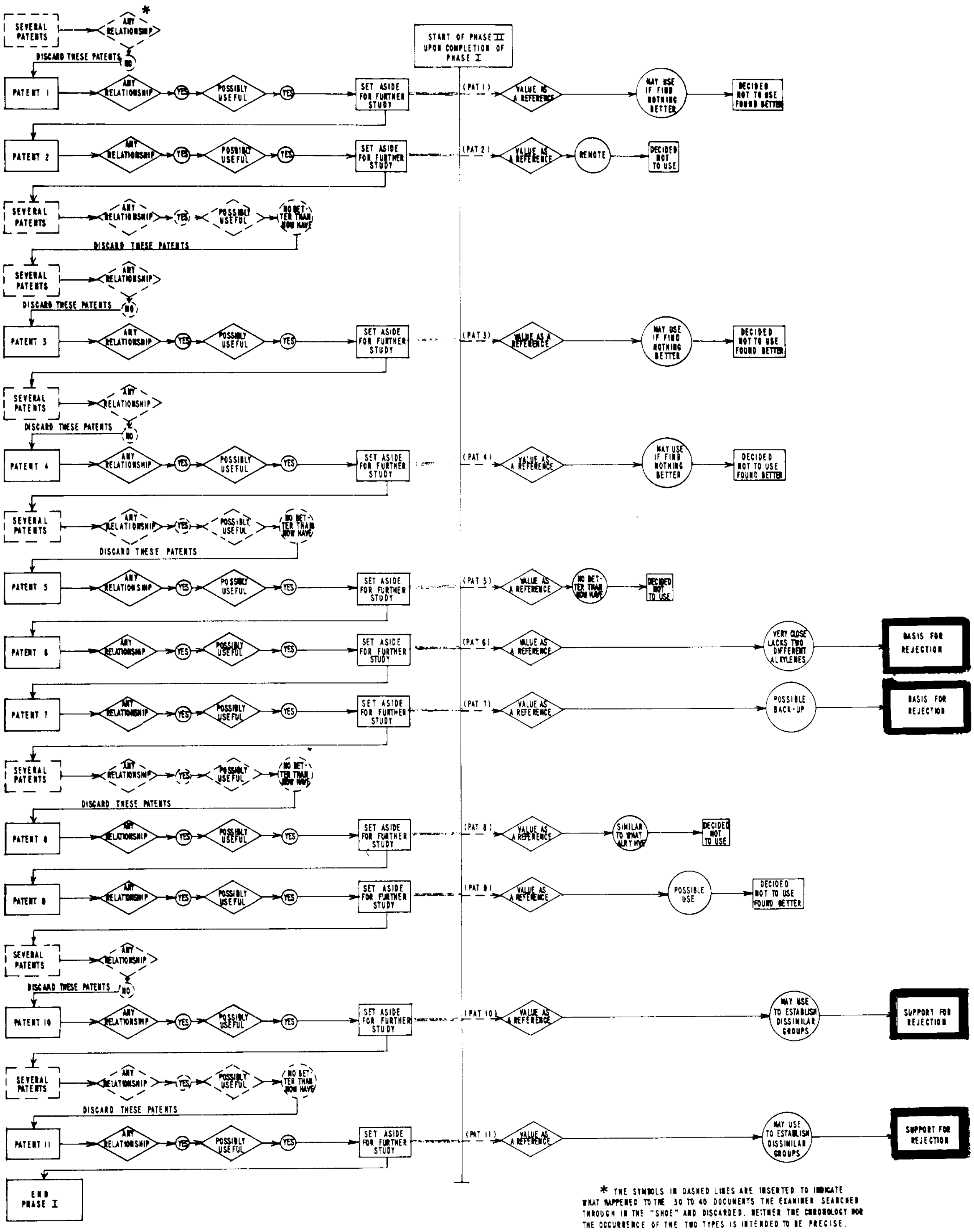

Figure 6. Schematic presentation of the sequence of activities in the manual search. The reader will observe that the search continued long past the finding of the two references used as bases for rejection of the application.

fication plan. When the examiner enters the manual classification system, he must first classify the information in the application correctly himself and then locate the most nearly similar existing file.

The mechanized systems are organized to make information retrievable on the basis of either concept or terms, so the mechanized file can be queried on many combinations within the principle–terminology continuum. The whole system is available at all times and inquiry can be pressed in several directions simultaneously. The examiner needs only be concerned with formulating as exact a statement of the principle and its satellite details as the system permits for the application at hand. He does not have to classify the concept or physically locate the subfile. The machine creates a subfile for him in response to his request.

3. *The examiner's approach to the acceptance of a reference as possibly pertinent seemed to be different in the manual and machine searches.* With machine search he apparently started out by looking for an "all or nothing" reference. He expected that, if he had phrased his questions properly, as nearly pat a reference as existed would be furnished ready for his inspection. He did not look actively for partially applicable documents, having in mind, probably, as he said, that previous use of this particular system had indicated such activity was unrewarding.

In the manual search, on the other hand, he accepted as his temporary standard the first patent he came across which seemed to have any degree of pertinence at all. He laid this one aside. Each patent he looked at from then on was examined in relation to the possibly useful reference he had already picked out: was the later one any better than the earlier find? If it was, he then accepted it as his new standard but all patents with apparent applicability were laid aside for a second, more careful screening.

It becomes evident in the restatement of the general search strategies that the *order* in which the patents were presented for inspection had considerable influence upon the ultimate choice of those considered to be most pertinent. One patent in the mechanized search (No. 15 on the chart) was thought to be a very close reference. In the manual search the same patent (No. 4) was discarded as being less applicable than others, yet one of those accepted as a very good reference in the manual search (No. 10) was read and discarded in the mechanized (No. 8). The order of presentation was evidently functioning here. The order effect seems to be important enough to warrant further study.

4. *In the manual search the examiner actually examined closely more documents than he did in the mechanized.* By continuing the manual search as far as he did, he found references which seemed more satisfying to him than those he would have been *able* to base a rejection on. He early found a reference that was very similar to the application except it was directed to a dialkylenetriamine compound rather than one in which an alkylene group was replaced by a cycloalkyl. He then turned to trying to plug this hole by finding a document which would say the unsymmetrical arrangement was obvious. He finally had to accept a statement of equivalence instead.

With the mechanized search he looked at only one more patent after having found what he thought was an adequate basis for rejection, even though he examined only 16 out of the 25 references on his primary list. With the manual search he continued looking until he had gone through all the patents in the digest, which he had designated as his primary searching unit. Figure 6 indicates that he read only five more than he absolutely needed, but this figure is not entirely accurate since we did not keep count of exactly how many patents were filed in the shoe.

5. *The periods of greatest examiner mental activity occurred at different times in the manual and machine searches.* In the manual procedure the examiner thought most actively during the actual search itself, the interval during which he scanned the prior art. He quickly determined relevance or no relevance of each document and then proceeded to look at the patent in the file. In the mechanized search, on the other hand, the greatest period of mental activity occurred during the planning of the search strategy and the formulation of the questions for the machine. At this time the examiner went through an intensive review of both his pertinent chemical knowledge and his understanding of the organization of the mechanical retrieval system.

Both manual and machine searches were characterized by a high level of mental activity during the period of second review of the art selected as having some relevance. Examination in the manual search was more prolonged and perhaps more thorough. The manual search had produced a larger number of prior art documents than the mechanized, and the examination did not stop as soon as citable references had been found.

6. *In psychological terms, the feeling tones of the protocols for mechanized and manual searches were different.* Both searches were tape recorded and played back later in close succession so the difference in treatment was evident. The mechanized search protocol produced a brisk and efficient tone, the kind that is characteristic of a person doing a well-mastered routine job. The work was there to be done; the examiner would do it quickly and get on. It should be added that the mechanized search was performed under much less pressure from other duties than was the later manual search but, in contrast, the tone of the manual search was anticipatory, pleased with the challenge of the problem, interested. The procedure evidently became a vital, personal endeavor.

With the mechanized search a feeling of defensiveness was apparent, as though he must defend his interpretation and use of the system against critics. He may have felt that he was being forced to employ methods devised by others, "unnatural" methods not of his own choosing or creation, and that his performance with that system would be inspected critically.

7. *The machine does not have "hunches."* The machine knows no short cuts. It does exactly what it is told, no more, no less. It is completely literal and accepts no vague orders to report out on incidental discoveries. The examiner must be the source of the hunches, and with mechanized search they do not seem to come as readily as with manual. The machine perhaps comes between the examiner and his thoughts in forming hunches. It can be hypothesized that better orientation about the characteristics of the system, or longer experience with it, might lead to more examiner insights while using it.

The most efficient and rewarding use of the machine naturally requires intimate knowledge of the system capabilities and performance. The imagination required for forming hunches must be more disciplined or channeled in order to conform to the machine system and structure. This happy combination of circumstances and capacities is achievable only after intensive training and long experience. As this examiner said:

"There is a multitude of areas that you can diversify or vary for combining references, but then it is virtually impossible to code questions, on the one hand dropping feature A and on the other broadening feature B, and run all these searches and then try to composite them so you don't look at the same document time and again or even go through all the documents that fall out."

8. *Mechanical search by this system did not appear to encourage browsing.* The documents which are not close references often bear little relationship to the concept under examination. Because the system was generated by placing the descriptors and all of the compounds for a patent on a single card, the machine "drops out" the numbers of patents that may use an essential fragment in an entirely different type of compound than that being sought. This process is known as "compositing." As the examiner himself put it:

"Your can't browse through "no drops" or false drops. (The machine either didn't locate any references or else listed irrelevant ones.) Sometimes one of them will give you a starting point but it isn't often. The false drops are much farther from the central idea of the search than are irrelevant patents in the same subclass."

He evidently did not expect "excursions" to be rewarding.

In manual search the examiner can approach the file with several alternate questions at the same time. If he tried to apply the same technique to searching by this mechanical system, it would take repeated machine queries, considerable clerical work, and a good deal of time. A more sophisticated or larger system would permit more varied simultaneous inquiries.

In order to "browse" by machine, either the questions asked must be broadly stated, so as to cover related areas, or many more narrowly based inquiries must be successively addressed to the machine to probe areas as the examiner thinks of them. The latter practice is time-consuming unless the questioner can work "on-line" to the stored information. With a mechanized system such as the one used by Mr. Levow, an examiner would have to *intend* to "browse". Incidental but pertinent information would not be likely to arise as long as the questioner pursued a policy of trying to locate a "pat" reference with the greatest possible dispatch.

The comparison between manual and machine search in the organometallics art, as presented in this paper, has concentrated on the basic difference in mental activities that occur with use of the two systems. Further work is needed to quantify and support the hypotheses presented here. The present analysis is to be considered as developmental only.

Substructure Searching of Chemical Compounds Using Polish-Type Notation

JOHN T. WELCH, Jr.
Goodyear Aerospace Corporation, Akron, Ohio

Received March 19, 1965

The utility of Polish notation for representing chemical compounds in a compact machine-readable form has been documented in this journal by Eisman (*1*) and Hiz (*2*). The familiar plane structural diagram may be considered a finite, connected, weighted, and, in many cases, undirected linear graph. (Graph-theoretical definitions of these terms need not necessarily concern the reader. Berge (*3*) and other basic texts are available, however.) A Polish string of symbols may then be written to completely represent any tree of the graph. By the term "substructure search," we shall mean the attempt to find a partial subgraph of the graph of the compound which matches node-for-node, *i.e.*, which is isomorphic to, the graph of the substructure. The very real limitations of these structural diagrams in representing chemical compounds will not be considered.

Eisman and Hiz are not concerned with the Polish notation as a medium for substructure search, and both show how the notation can be internally converted to other forms of representation. It is the purpose here to demonstrate the potential of the notation as an internal symbolic language for substructure search, and to indicate the manipulative and bookkeeping abilities which it brings to this task.

The particular variation of the Polish notation to be used will be described very briefly in the next section. Following that, and a general outline of the strategy to