DOCUMENT RESUME

ED 110 016 IR 002 284

AUTHOR	Zunde, Pranas
TITLE	Scientific and Technical Information Transfer for Education (STITE). Research Report No. 4.
INSTITUTION	Georgia Inst. of Tech., Atlanta. School of Information and Computer Science.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	Jun 75
NOTE	159p.; For related documents see ED 095 867 through 869
EDRS PRICE	MF-$0.76 HC-$8.24 PLUS POSTAGE
DESCRIPTORS	Annual Reports; *Computer Programs; Concept Formation; Data Bases; *Experimental Programs; Information Centers; *Information Dissemination; Information Retrieval; Information Scientists; Information Storage; Information Systems; Instructional Design; Learning; *Science Instruction; Science Materials; *Science Teachers; Scientific Concepts; Use Studies
IDENTIFIERS	*Project STITE

ABSTRACT

Emphasizing the design of a data base management system for the experimental STITE (Scientific and Technical Information Transfer for Education) project, this progress report details the emerging features of this projected facility. Compiled by four STITE researchers, the report examines: science information communication, learning, dissemination, and the structure of STITE itself. In addition, 76 flowcharts are presented which document the STITE data base management system. (DS)

RESEARCH REPORT

SCIENTIFIC AND TECHNICAL INFORMATION TRANSFER FOR EDUCATION

(STITE)

REPORT NO. 4

Pranas Zunde
Project Director

June, 1975

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA

RESEARCH REPORT

SCIENTIFIC AND TECHNICAL INFORMATION TRANSFER FOR EDUCATION

(STITE)

Pranas Zunde
Project Director

June, 1975

GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA

SCIENTIFIC AND TECHNICAL INFORMATION TRANSFER FOR EDUCATION

(STITE)

Fourth progress report on research performed at the School of Information and Computer Science, Georgia Institute of Technology, Atlanta, Georgia, under National Science Foundation Grant No. GN-36114.

Pranas Zunde
Project Director

June, 1975

PREFACE

During the period of research work described in this report, main emphasis was placed on the design of a data base management system as part of the projected experimental STITE facility. Some of the emerging features of this system are described in detail in Part Four of the report. Parallel to this development, further studies were made both for the purpose of sharpening certain conceptual prerequisites for the design of the STITE interface system and for obtaining further insights into potential user needs and requirements. These studies are described in Parts One, Two, and Three.

Participation in this phase of the study of Dr. L. J. Gallaher, Senior Research Scientist, Dr. T. C. Ting, Associate Professor of Information and Computer Science, Dr. Albert N. Badre, Assistant Professor of Information and Computer Science, and Mrs. Dorothy S. Hughes, Research Analyst and Librarian, is gratefully acknowledged. They appear as authors or co-authors of the portions of this report to which they made major contributions.

TABLE OF CONTENTS

TABLE OF CONTENTS

(continued)

APPENDIX

PART ONE

ON THE STRUCTURED COMMUNICATION OF SCIENTIFIC AND TECHNICAL CONCEPTS

PART ONE

ON THE STRUCTURED COMMUNICATION OF SCIENTIFIC AND TECHNICAL CONCEPTS

Albert N. Badre

I. INTRODUCTION

One of the goals of the STITE project is to analyze and define the process by which science learning materials are developed and structured. This task would require the intuitive identification and definition of the elements of science learning which are used in the instructional communication of scientific and technological concepts. One of the key questions that might be considered in this context has to do with the elements associated with the design and structure of the "course". In an earlier STITE report, it was suggested that the "course", or the "science-subject space", consists of a number of major elements associated with modes of presenting the learning material such as illustrative narrative presentations, definitional narrative presentations, problems, examples and exercises, conjectures and hypotheses (9).

The present document aims at analyzing the elements of instruction in terms of the process of developing a science-subject space. This will be done by considering two main questions:

(1) What are the prerequisite concept-communication units of an effective verbal-instructional presentation?

(2) How can the specification of levels at which a scientific concept is communicated lead to a precise classification of user course-development characteristics for the STITE system?

These are two questions which should not be disregarded in the process of developing a course. Accordingly, they must be taken into account in any attempt to design an artificial system which is intended to aid the instructor in the transfer and utilization of scientific and technical information from its present repositories to the instructional milieu.

The first question is motivated by the recognition that there are certain basic factors associated with verbal concept communication which are independent of the subject-matter or the instructor per se and which impose necessary conditions required by the effort to make the instructional delivery effective. The second question follows from the assumption that there are certain goal-conditions associated with user-instructor characteristics and imposed on the communication process by the individual instructor which can be classified in terms of instructional objectives.

II. THE CONCEPT IN EFFECTIVE INSTRUCTIONAL COMMUNICATION

It may be suggested that the basic unit of instructional communication is the concept. Initially, a concept may be defined as the sum total of the intensional attributes of the object of reference along with an associated set of relations over the attributes.

Given that the proper management of science concept learning in an instructional setting provides for a generally more adequate attainment of the overall learning objectives, then the problem of developing instructional programs such that a learner's <u>concept attainment</u> is maximized becomes one of fundamental concern for the design of an instructional system or the design of a system that aids the instructor in the processing and structuring of information for learning situations.

III. CONCEPT ACQUISITION

It seems appropriate to start by analyzing the concept-unit of information in terms of the processes of acquisition and learning. Concept learning is recognized in terms of two basic processes: *Concept formation* and *concept attainment* (3). The process of concept formation involves the learning of new concepts. It is one whereby a learner acquires an extensional recognition of the concept at hand. That is to say, when a learner encounters a totally new stimulus and simultaneously (or subsequently) comes into possession (whether through learning or assignment) of the name of the concept of which the newly encountered stimulus is a positive instance, then it can be said that he has "formed" the given concept.

Concept attainment, on the other hand, requires an intensional understanding of the concept to be learned. That is, a learner is said to have "attained" a concept, if, after he has already "formed" it, he is able to identify the same concept in terms of its necessary and sufficient attributes or dimensions. The notion of "attainment" requires that the learner, given a class of instances, be able to *discriminate* between those which do not belong to the concept and those that do on the basis of his knowledge of the concept's necessary and sufficient attributes. His intensional understanding of the concept will presumably enable him to eliminate those attributes which are irrelevant.

As an illustration of the above, suppose that in a course on graph theory, a learner encounters for the first time an instance of the concept which he is told is called "graph". By thus encountering an instance of the concept and labelling it "graph", he is said to have *formed* the concept of "graph". Then, once he is able to discriminate "objects" which are called "graphs" from those which are not, on the basis of the relevancy and irrelevancy of the attributes of the concept, "graph", then it can be legitimately said that he has *attained* the concept of a "graph". Moreover, it is important to note that concepts become more complex as the number of attributes and their values increase. For example, if one learns that there are different kinds of graphs, e.g. Eulerian graphs, Hamiltonian graphs, then it is not only necessary to learn how to distinguish between graphs and non-graphs, but also to be able to recognize and discriminate between at least two sub-categories of graphs.

Hunt (6) suggests that the learning of concepts occurs only at the formation stage. That is, the process of concept formation is coterminous with that of concept learning. He views the attainment stage as merely being the stage of identification of an already learned concept. Thus, to Hunt, concept learning occurs at the formation stage where one presumably learns a *new* concept. Once that concept has been assimilated into the learner's cognitive structure, it becomes the object of retrieval and recoding. It becomes a problem for semantic memory.

This distinction made by Hunt seems not so much to contradict as to clarify and make more useful, at least from a processing approach, the Bruner categorizations. It is appropriate to be able to distinguish in the framework of instructional communication new concepts not yet learned, and concepts already learned but which fall under some other category of concept learning, e.g. concepts already learned but not mastered, ambiguous concepts that need to be reduced to the relevant meaning, and misconstrued concepts that require re-learning (1). This categorization, in turn, leads to some useful generalizations about concepts learned in an instructional setting.

IV. CONDITIONS IMPOSED BY THE INSTRUCTIONAL COMMUNICATION OF VERBAL CONCEPTS.

There are several conditions which have been empirically shown to be associated with an instructional event when such an event involves the communication of verbal concepts (1). Concept acquisition in instructional communication can be viewed as consisting of associating a set of attributes with the name of a concept. An attribute can be viewed as that [illegible] on without which a given entity will no longer be precisely that given entity. Moreover, concepts can be thought of operationally as the mental abstractions of life-experiences in which an experience is defined as the response to any stimulation. This would lead to the concept that learning is directly influenced by the reinforcement given a response. Moreover, the learning of a concept is strengthened if the experiences of the learner in relation to the concept are in some respect similar. The way in which those experiences are similar is in a sense an expression of the concept. In this sense, the experiences are the positive instances of the concept. Another necessary condition is that a set of negative instances of the concept be generated and stored in the repertoire of the learner's experiences. Still another necessary condition is that there be an appropriate *sequencing* of positive and negative instances. This sequencing is important because the complexity of a concept increases as a function of an increase in the number of relevant attributes. In addition the number of values that each attribute has may increase, (e.g. man; black man; tall black man; tall and strong black man; tall, handsome, and strong black African man). Hence concept attainment will necessarily become more difficult. Thus, the sequencing, cognitive mapping, and mental organizing of attributes and negative instances of a concept become highly essential. Still another necessary condition is that a concept be presented to the learner in as many different classes of tasks as possible. Johnson and Stratton (7) report that when, in an experiment, they classified a concept, defined it, used it in a sentence, and gave it synonyms, they produced superior results to those produced when only one of these four m[illegible] was used to explicate the concept.

This leads to the description of a concept acquisition model which was developed previously (1) in order to describe and use a scale for verbal concept acquisition. This model assumes that the concept-unit of instructional communication is a function of the various combinations of relevant and irrelevant attributes.

> "Let ZAID denote the name of a concept, and Z.A.I.D. be the conjunction of its necessary and sufficient attributes; $Z'.A'.I'.D'.$, the irrelevant attributes of ZAID; $\bar{Z}.\bar{A}.\bar{I}.\bar{D}.$, any set of attributes not related to ZAID; and z.a.i.d. another set of necessary and sufficient attributes of ZAID, where z.a.i.d. $\neq$ Z.A.I.D.
>
> Response categories. Then, if learner's response to what is ZAID is: (a) Z.A.I.D., concept is attained; (b) none of the above classifications, concept is new; (c) Z.A., Z.I., Z.D., such that if and only if a subpart of the conjunction of necessary and sufficient attributes is given, concept is vague; (d) $Z.\bar{Z}.A'.$, $\bar{Z}.\bar{A}.$, $Z'.A'.D'.$, $Z'.A'.I'.D'.$, $Z'.A.I.$, such that the response entails either irrelevant or nonrelevant attributes or both, concept is misconstrued; (e) z.a.i.d., and only z.a.i.d., concept is ambiguous; (f) z.a.d. = ambiguous-vague intersection; (g) $z.a.\bar{Z}.$ = ambiguous-misconstrued intersection. (1)

It seems that when the formation-attainment (learning-identification) continuum is considered in terms of its pedagogical significance, two fundamental assumptions must be kept in mind.

First, that whenever the learning of a concept occurs in an instructional setting, it generally is contextually-oriented. For example, it would not be useful to introduce the mathematical concept of a "derivative" into an instructional dialogue about the merits or demerits of peaceful coexistence. As

Carroll (4) suggests, it would not even be useful to bring up such a concept in a mathematical dialogue in which, for example, the learner was not familiar with the prerequisite concepts of "algebraic function" and "slope". Hence, concept learning in pedagogy takes place in a definite context. Therefore in developing a course that requires the communication of concepts, as would be the case with courses developed from STITE-type information, the instructor would have to utilize techniques which place the concept in the proper context.

Secondly, it is assumed that in any pedagogical event where a concept is to be communicated, the learner's conceptual state of mind immediately prior to the communication of the concept falls somewhere on the formation-attainment continuum or on a variation thereof. This assumption implies that in the communicating of concepts, full regard must be given to individual differences. That is to say, cognizance must be taken of the individual's level of "mental" relationship or preparedness to the concept at hand.

Both of these assumptions are well underlined by Gagne's (5) conditions of learning model. Gagne has argued and others have shown that the acquisition of a skill (in this case, one that requires cognitive processing) depends on and perhaps contains the knowledge and mastery of other prerequisite skills. He classifies skills in learning into eight types of performance that must be mastered from the simple to the more complex levels of cognitive activity. Kemp (8) summarizes the eight levels as follows:

1. Signal learning
2. Stimulus-response learning
3. Chaining
4. Verbal association

5. Multiple discrimination
6. Concept learning
7. Principle learning
8. Problem solving

Science teaching or the instructional communication of scientific concepts is more likely to be associated with the last four levels of cognitive skill acquisition. Thus, the learner's cognitive preparedness in terms of the last four categories would have to be taken into account as one is organizing and sequencing a course or a "science subject-space".

V. VARIATIONS IN USER CHARACTERISTICS IN TERMS OF INSTRUCTIONAL OBJECTIVES

The set of conditions listed above constitute those that are necessary to the instructional communication of science information and thus impose vital constraints on the instructor's activity in developing a science subject space. Because of the claimed universality of these conditions over science information transfer in learning, it becomes advantageous to specify them as a set of rules and algorithms with respect to a chosen topic and then to program the specifications as part of the STITE system. An example would be the development of a sub-course on the "topic", Eulerian graphs, as presented in an earlier STITE report (9). In this regard the system becomes an aid to the instructor in that if, in developing a course, he should violate one of these universal conditions, the program will interrupt with a violation note. By the same token, there are a certain group of conditions, such as learning objectives, which the instructor may impose on the instructional event when developing a course. To this extent, course-development becomes a collaborative effort between STITE and the instructor. The final product, the course, will as such be guided by a combination of necessary and optional conditions. The optional conditions, those associated with the learning objectives, will be instructor-specific.

Three categories of objectives have been identified in the literature (9). These are: (a) the cognitive domain objectives, e.g. naming, solving, remembering, proving; (b) the affective domain objectives, e.g. enjoying, appreciating, respecting, and (c) the psychomotor objectives, e.g. constructing, adjusting, manipulating. It is evident that the instructional objectives of courses developed with STITE-type information would primarily belong to the cognitive domain. Under the cognitive domain, there are several levels at which an instructor may choose to communicate his topic:

(1) He could specify that the communication be at the comprehension level, in which case terms such as "explaining by defining", "giving examples", or "summarizing" would apply.

(2) Another level of instructional communication would be that of application where terms such as using, solving, and proving would be proper.

(3) The instructor could choose his objectives at the analysis level in which case the terminal behaviors would be ones such as inferring and relating.

(4) The instructor could develop his course with the objective of achieving synthesis where organizing and planning may be the desired objectives.

By allowing for flexibility in the level of objectives, STITE can be viewed as a collaborative, adaptive system that allows for a wide range of user's objective-specifying characteristics. Thus it would combine the necessary conditions that accompany the instructional communication of STITE-type information with the flexible ones that accompany the wide variety of user-types.

VI. REFERENCES

1. Badre, A. N., and D. D. Starks. "Concept Learning As a Function of Ability and Method." Proceedings of the American Psychological Association 8:669-670 (1973).

2. Bloom, Benjamin, ed. Taxonomy of Educational Objectives; the Classification of Educational Goals. New York, D. McKay, 1956. 2 vols.

3. Bruner, J. S., J. J. Goodnow, and A. A. Austin. Study of Thinking. New York, Wiley, 1956. 330 pp.

4. Carroll, John B. "Words, Meanings, and Concepts." Harvard Educational Review 34:178-202 (1964).

5. Gagne, R. M. The Conditions of Learning. New York, Holt, Rinehart, and Winston, 1970. 407 pp.

6. Hunt, E. B. Concept Learning. New York, Wiley, 1962. 286 pp.

7. Johnson, D. M., and R. P. Stratton. "Evaluation of Five Methods of Teaching Concepts." Journal of Educational Psychology 57:48-53 (1966).

8. Kemp, Jerrald E. Instructional Design. Belmont, California, Fearon Publishers, 1971. 130 pp.

9. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia School of Information and Computer Science, Georgia Institute of Technology, July, 1973. Progress Report, NSF Grant No. GN-36114. 169 pp.

PART TWO

SCIENCE INFORMATION TRANSFER FOR LEARNING

PART TWO

SCIENCE INFORMATION TRANSFER FOR LEARNING *

Albert N. Badre, Dorothy S. Hughes, T. C. Ting, and Pranas Zunde

I. INTRODUCTION

One result of research and development in the field of science information in recent years has been the establishment of large banks of descriptive information and bibliographic data that is stored on digital and analog media. These collections of data, along with the mechanisms for their organization, search, and dissemination, comprise science and technical information systems.

In the past these centers have been mainly used by industry and research. There were indications that the use of such information systems for educational purposes was very limited. The results of the present study support this generalization. The main objective of the study was to ascertain the actual extent of usage of these centers by various categories of science educators. It is further hoped that this study will help to determine under what conditions and for what reasons science educators might make more intensive use of science and technical information centers. The work reported here is part of an ongoing study on ways and means of enhancing the transfer of scientific and technical information from its present repositories to science education systems (8).

---

* This is the final revised report. An earlier paper, published in the Proceedings of the 1974 ASIS national conference, gave prelimi results of the study based on partial, or 22.5%, returns.
The present report is based on a completed total of 33.1% returns o
2000 questionnaires sent.

II. PREVIOUS USER STUDIES

The information explosion has been accompanied by an exponential growth of studies on information users and their needs. However, in the field of science, the bulk of studies conducted has dealt mainly with the information needs of scientists in general and few have compared information needs between disciplines. Among existing studies there are relatively few that include findings in the area of science education, and these usually have been comparisons of information needs based on the activities of the scientist, i.e. whether he is engaged in research, teaching, or industry. An example of a comparison study in which some data about science educators may be found is an early investigation by Törnudd (7).

Even more unusual are user studies that are directed specifically and explicitly towards the needs of science educators. Menzel, for example, studied the information exchanging behavior of a group of university teaching scientists (6). His conclusions were made with respect to a generalized concept of the teaching scientist and did not take into account the relationship between various scientific disciplines and information acquisition behavior.

Other investigators have taken a broad sub-category of science, such as social science, physical science, or engineering, and have attempted to define user needs with respect to these broad categories (1) (2) (5). Many of these findings have failed to delineate information needs that are unique to the particular discipline and, furthermore, supporting data that would allow the extension of findings to other disciplines was not presented. For example, Bartkus concludes that the engineering educator is concerned with the needs, among other, of "his research students to assure coverage of previous work." (1) It cannot be claimed that this finding is unique to engineering educators but, based on the data collected, no valid generalizations to other disciplines can be made.

Also directly related to the present research are those studies concerned with educator requirements in using information systems. Most of these, of which there were few to begin with, have not been directly concerned with scientific and technical information centers. For example, Baughman investigated the information needs relative to the more general category of educational information centers, while Borman and Mittman examined the behavior of users of new information systems, but neither study was information center-specific nor discipline-specific (3) (4).

III. GOALS OF THE PRESENT STUDY

While previous studies reveal some useful findings about the information needs and acquisition practices of science educators, there are two important factors that have received little emphasis.

(1) None of the above types of user studies has dealt specifically with the needs of science teachers who are potential or actual users of science and technical information systems;

(2) All of the above cited studies have dealt with the information needs of educators in general. There has been no attempt to investigate those needs, except in a very limited way, by academic disciplines.

The present study was intended to fill this gap to some extent. In particular, the goal was to obtain data relative to the following hypotheses:

a) A large percentage of science educators are not aware of the availability or of the existence of science and technical information systems.

b) Most science educators who are aware of the availability of science and technical systems have no ready or easy access to them.

c) Although most academic scientists are not aware of scientific and technical information systems, those who access such systems do so for research, and not for teaching purposes.

d) Most of the science educators who have access to such systems find that the access tools are inflexible and unsatisfactory to use.

e) Most science educators who access information from science and technical information centers find the information they receive to be of little use for their instructional purposes.

IV. METHOD OF INVESTIGATION

Since it was important to access a large segment of science educators in colleges, it was decided that the most appropriate technique for the purposes of this study would be the mailed questionnaire. Accompanying each questionnaire was a letter to the potential respondent explaining the purposes of the study and the meaning of scientific and technical information systems. The questionnaire was relatively short; it was estimated, on the basis of pre-testing, that a potential respondent would take no more than ten minutes to complete it.

The sequence of questions was designed to be logically ordered. In organizing the questions, the respondents were divided into two categories: (a) Those who use science and technical information systems, and (b) Those who do not use such systems. This division permitted three categories of questions based on the type of respondent. The first category of questions was directed only to those who actually use science and technical information systems. The second category was directed to those who do not use such systems. The last category of questions was directed to both types of respondents, users and non-users.

In total, the questionnaire was made up of eleven questions with instructions on what to do in order to answer each of the questions.

A stratified-random sampling technique was used in selecting approximately 2,000 science educators from colleges and universities in the United States as potential respondents. The names of the educational institutions were obtained from American Universities and Colleges, 1973 edition, and the individual faculty names were obtained from general catalogs of the colleges and universities selected. More specifically, the sample was constructed as follows:

1. Within the three major scientific divisions, namely the biological, the physical, and the social sciences, both the traditional disciplines, as well as inter-disciplinary programs under each one of these broad science categories, were considered. In the biological sciences, the following fields were selected: biochemistry, botany, genetics, microbiology, physiology, zoology; in physical sciences: aeronautical engineering, astronomy, chemistry, civil engineering, electrical engineering, mechanical engineering, statistics, physics; in the social sciences: anthropology, economics, library science, psychology, sociology.

2. Within each field, colleges and universities that offer such a discipline were identified by listings in American Universities and Colleges. Approximately 108 institutions were randomly selected from the list.

3. Between five and ten faculty names were selected from the catelogs of the selected institutions relative to the chosen department or program. The selection of individual names was at random. Consequently, the proportion of faculty ranks can be expected to reflect the actual composition of the faculty population in the field. A total of approximately 100 faculty members were selected from each chosen field to receive the questionnaire.

V. RESULTS

It was considered that, given a total of 2000 questionnaires, a response size of 15% to 20% would be sufficient to allow meaningful interpretation of the data. The sample size on which this analysis is based represents a 33.1% response rate.

TABLE I LEVEL OF AWARENESS OF INFORMATION CENTERS

a. PHYSICAL SCIENCES

	Yes - %	No - %
Aeronautics	61.5	38.5
Astronomy	45.8	54.2
Chemistry	50	50
Civil Engineering	51.6	48.4
Electrical Engineering	48.2	51.7
Geology	37.5	62.5
Mechanical Engineering	52.3	47
Physics	23.3	76.6
Statistics	11.5	88.5
TOTAL	42.6	57.4

b. BIOLOGICAL SCIENCES

	Yes - %	No - %
Biology	38.2	61.7
Botany	47.6	52.4
Physiology	52.1	47.8
Zoology	38.8	61.2
TOTAL	42.4	57.6

c. SOCIAL SCIENCES

	Yes - %	No - %
Anthropology	21.7	78.3
Economics	8	92
Library Science	43.5	56.5
Psychology	45.4	54.5
Sociology	30.8	69.2
TOTAL	27.1	72.9

	Yes - %	No - %
d. OVERALL TOTAL	39.2	60.8

The findings with respect to the first hypothesis, that most science educators are not aware of scientific and technical information centers, are outlined in Tables Ia, Ib, and Ic. The results show that in all three categories, the number of science educators who are not aware of such centers is greater than those who are. However, the level of awareness varies over the three major science divisions. It can be seen that the physical sciences have a much higher level of awareness than the social sciences. Chemists, astronomers, civil engineers, and physiologists seem to be among those who are most aware. These findings, combined with those of Table II seem to confirm the assertion that science and technical information centers are publicized mainly as services for the research needs of physical and biological sciences. Psychology is the only discipline in the social sciences that seems to have a high level of awareness. This may be due to the fact that psychology already has an existing specialized information service, e.g. *Psychological Abstracts*, while the other social sciences are not so well covered. In addition, psychology has less defined boundaries, thus allowing its scientists more contact with the other scientific divisions than is true of the other social sciences.

TABLE II RESEARCH VS. TEACHING USAGE OF INFORMATION CENTERS

a. PHYSICAL SCIENCES

	Research %	Teaching %
Aeronautics	85.7	14.3
Astronomy	88.8	11.1
Chemistry	68.3	31.7
Civil Engineering	66.6	33.3
Electrical Engineering	92.9	7.1
Geology	75	25
Mechanical Engineering	80	20
Physics	100	0
Statistics	100	0
TOTAL	77	23

b. BIOLOGICAL SCIENCES

	Research %	Teaching %
Biology	75.9	24.1
Botany	83.5	16.7
Physiology	78.5	21.4
Zoology	81.3	18.7
TOTAL	79.2	20.8

c. SOCIAL SCIENCES

	Research %	Teaching %
Anthropology	71.4	28.6
Economics	100	0
Library Science	63.6	36.3
Psychology	70	30
Sociology	77.7	22.2
TOTAL	71	29

	Research %	Teaching %
OVERALL TOTAL	76.8	23.2

In order to determine the relationship between awareness and use, the data in Table III was collected. It indicates the relationship between those who are aware of information center services and those who actually use those services. It seems obvious from these results that if the scientist is aware of information centers, there is a high likelihood that he will use their services.

TABLE III RELATIONSHIP BETWEEN AWARENESS AND USE OVER THE THREE MAJOR DIVISIONS

	Yes - %	No - %
Physical Sciences	78.3	22.7
Biological Sciences	82.2	17.8
Social Sciences	70	30
TOTAL	76.9	23.1

The second hypothesis, that ready and easy access to scientific and technical information centers is not available, was tested by eliciting information as to whether scientists use those services directly or through other channels. As Table IVa, IVb, IVc, and IVd show, approximately 50% accessed the services through indirect channels such as libraries.

It was hypothesized that most academic scientists who access information centers use such services for research and not for teaching. The results shown in Tables IIa, IIb, IIc and IId indicate that the primary use of information received from centers is for research. This was an expected finding since the science and technical information centers were designed to serve the information needs of research scientists. There were no noticeable differences between the three major divisions of science nor between individual disciplines.

TABLE IV DIRECT AND INDIRECT ACCESS TO INFORMATION CENTERS

a. PHYSICAL SCIENCES

	Direct - %	Other Channels %
Aeronautics	42.8	57.1
Astronomy	53.8	46.1
Chemistry	51.9	48.1
Civil Engineering	62.5	37.5
Electrical Engineering	42.8	57.1
Geology	0	100
Mechanical Engineering	44.4	55.5
Physics	37.5	62.5
Statistics	50	50
TOTAL	51	49

b. BIOLOGICAL SCIENCES

	Direct - %	Other Channels %
Biology	57.6	42.3
Botany	41.2	58.8
Physiology	54.5	45.4
Zoology	50	50
TOTAL	52.9	47.1

c. SOCIAL SCIENCES

	Direct - %	Other Channels %
Anthropology	33.3	66.6
Economics	0	100
Library Science	71.4	28.5
Psychology	62.5	37.5
Sociology	57.1	42.8
TOTAL	55.2	44.8

	Direct - %	Other Channels %
d. OVERALL TOTAL	51.5	48.5

Tables V and VI reflect the findings with respect to the fourth and fifth hypothesis concerning the levels of satisfaction with the services of information centers and the relationship between the information requested and the information received. It can be seen that those who are regularly or often satisfied with the services of information centers constitute more than half of the users represented. However, it must be noted here that, with respect to the fifth hypothesis, the findings reflected in Table VI must be considered in relation to the findings in Table II. Since approximately 80% of all users accessed centers for research purposes, it becomes clear that the high rate of correspondence between material request and that obtained applied primarily to research, and less to the instructional needs of the scientist. From this finding, it appears that the science educator, regardless of his subject field, goes elsewhere for information for his instructional purposes.

TABLE V LEVEL OF SATISFACTION WITH INFORMATION CENTERS

	Never-%	Sometimes-%	Often-%	Always-%
Physical Science	1	24.2	26.4	49.3
Biological Science	1.6	15.9	25.4	57.1
Social Science	3.5	7	24.1	65.5
TOTAL	1.6	18	24.7	55.7

TABLE VI CORRESPONDENCE BETWEEN WHAT IS REQUESTED AND WHAT IS OBTAINED

	Never-%	Sometimes-%	Often-%	Always-%
Physical Science	3.5	22.1	34.9	39.5
Biological Science	3.3	11.5	36.1	50.1
Social Science	0	12	36	52
TOTAL	2.9	16.8	35.5	44.8

VI. CONCLUSIONS

The results of the study support the two statements that the primary use of information centers is for research and that a majority of science educators are not aware of the services of information centers. However, most of the science educators who are aware of the services use the centers frequently and with satisfaction. It may be concluded that exposure of information services to science educators is an important factor in the improvement of the use of information centers.

VII. REFERENCES

1. Bartkus, E. P. "Major Sources of Information for Engineering Educators." Engineering Education, 60:377-380 (January, 1970).

2. Bath University of Technology. Information Requirements of Researchers in the Social Sciences. Bath, England, Bath University of Technology, 1971. 2 vols. ED054806 and ED054807.

3. Baughman, Robyn C. Survey of Information Needs of Educational Information Specialists. College Park, Maryland, University of Maryland, School of Library and Information Services, 1972. 31 pp. ED068101.

4. Borman, Lorraine: Mittman, Benjamin. "Interactive Search of Bibliographic Data Bases in an Academic Environment". Journal of the American Society for Information Science, 2:164-171 (May - June, 1972).

5. Brittain, J. M. Information and Its Users; A Review With Special References to the Social Sciences. New York, John Wiley, 1971, 208 pp.

6. Menzel, Herbert. "Planned and Unplanned Scientific Communication." In Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16 - 21, 1958. Washington, D.C. National Academy of Science - National Research Council, 1959. Vol. I, p. 199-243.

7. Törnudd, Elin. "Study of The Use of Scientific Literature and Reference Services by Scandinavian Scientists and Engineers Engaged in Research and Development." In Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16 - 21, 1958. Washington, D.C., National Academy of Science - National Research Council, 1959. Vol. I, p. 9-65.

8. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, June, 1973. Progress Report, NSF Grant No. GN-36114.

9. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, December 1973. Progress Report, NSF Grant No. GN-36114.

10. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, July, 1974. Progress Report, NSF Grant No. GN-36114.

PART THREE

ANALYSIS OF THE EDUCATIONAL USE OF AN INFORMATION CENTER

PART THREE

ANALYSIS OF THE EDUCATIONAL USE OF AN INFORMATION CENTER

Dorothy S. Hughes

I. INTRODUCTION

Results of the survey of the educational needs of science educators seemed to indicate that further investigation into the instructional usage of a specific information center might yield some helpful results, and a group of users of the Georgia Information Dissemination Center was selected by STITE for this inquiry.

The Georgia Information Dissemination Center, or GIDC, is a bibliographic retrieval system serving the faculty, research staff, and graduate students of the University System of Georgia. It began in 1968 and searches multiple data bases to provide both SDI and retrospective search services to its users. Physical facilities are located on the university campus at Athens, and remote users from Georgia's eleven other senior colleges and twelve junior colleges are usually served through reference librarians on those campuses.

A list of users of the GIDC was obtained from the Center. This user list contains approximately 4,000 names, along with the academic department with which the user is associated and his mailing address.

II. PERSONAL INQUIRY TO USERS

As a beginning, telephone calls were made to sixty-seven randomly selected users at the Georgia Institute of Technology and at Georgia State University to inquire if any of the information they had received from GIDC had been used for instructional purposes. When nineteen of those professors contacted indicated that such had been the case, then personal interviews were arranged with nine of them in five different departments of the two schools to explore at greater length the instructional usage that these professors had made of the information they had received. In addition, five of the telephone calls developed into interviews when the initial call aroused interest and subsequently extended the conversation into greater depth, so a total of fourteen personal interviews were conducted.

Each professor was asked in the interview how he became aware of the GIDC, the actual educational task that was accomplished with the information he received, if the information was directly usable as it was received in the print-out, and the features of an information system that would make it a better one for instructional purposes.

Awareness of the GIDC as it was indicated in the interviews came through formal and informal channels, with the informal ones seeming to be the stronger. In one case a memo and a brochure explaining the service was sent to the professor from the campus library, and in another a presentation about the service was made to the department by an information specialist from the GIDC. Two professors indicated that the deans of their departments had explained the service to them while two others indicated that they learned of it through colleagues. A student informed one professor that the GIDC was available, and another overheard a conversation about it between the dean of the department and a student. Finally, at least one professor just could not remember how he learned about it.

Actual educational tasks that utilized the information received from GIDC included course preparation, course up-date, preparation and up-date of reading lists and bibliographies, development of notes for a course for which no text existed, assistance in writing text books, and term paper assignments.

In only three instances was the information used directly from the print-out. One professor divides the print-out and distributes portions of it to his graduate students with instructions for them to locate, read, and annotate articles. This practice teaches the literature of the field and results in an annotated bibliography. A chemistry professor binds the print-outs in a looseleaf notebook and makes them available to students who are working in the area of the search. Another chemistry professor selects five or six references from the print-out to make subject lists which he then gives to students with instructions to locate and read the articles and then develop a term paper.

In most cases, however, the references from the print-out are retrieved and scanned or read before the information is utilized in any way.

Most of the suggestions for modifications in the GIDC to increase or improve its use by educators were concerned with practical aspects of its use rather than with the nature of the information itself. While several professors expressed the desire for the inclusion of abstracts at no charge and for a capability to retrieve examples and problems, most of them stressed simplicity and ease of use through more convenient methods of instigating searches, more accessible assistance in profile preparation and revision, and changes in the print-out which would increase its readability and ease of handling.

III. THE QUESTIONNAIRE TO GIDC USERS

Following the personal interviews with GIDC users, a further inquiry, in the form of a questionnaire, was sent to a larger sample of users.

This questionnaire (See Appendix B) was designed to determine what, if any, instructional usage the recipient makes of the information he receives from GIDC and his suggestions for change in the system that would be beneficial to the educator community.

Testing of the questionnaire was done with six professors on the Georgia Tech campus. After some discussion with those recipients and with Ms. Margaret K. Park of the GIDC, some revisions were made, and copies of the questionnaire in its final form were mailed with an explanatory letter and a return envelope to 1211 persons on the user list. (See Appendix A and Appendix B).

Of the 1211 sent, 410, or almost 33%, were answered and returned. Every respondent did not necessarily answer every question posed, but the following summary is based on the answers received. For example, many users did not indicate the subject area in which the information was used (Question 4), so tallies have been made from the answers received, and some reservations about general assumptions must be kept in mind.

Of the 410 users who responded, 251, or 61.2%, are apparently actively engaged in instructional activities as instructors, assistant professors, associate professors, and professors. (The category of "other academic" includes librarians, teaching assistants, research associates, directors, and information scientists while the non-academic category takes in administrators, project directors, program coordinators, managers, and those engaged in research only.) One hundred nineteen users were graduate students who may or may not be engaged in teaching.

One hundred eighty-eight persons, or 45.8%, indicated that the material they received from GIDC had been used for some type of instructional purpose, and Question 3 brought out the fact that the most common instructional use is the compilation of bibliographies or reading lists. This type of use was indicated by 137 users, or 33.4%. Utilization for current awareness in the subject area of a course was indicated by 98 respondents (23.9%), while 80 (19.5%) said that they used the information for collection of data. Updating an existing course was a utilization cited by 62 (15.1%), while the tasks of preparation of illustrative examples and development of new courses were indicated by 42 (10.2%) and 34 (8.2%) users respectively. The "other" category checked by 34 persons most often referred to directing research projects of students and developing research papers. Selection of case studies and preparation of quizzes and tests were less frequent uses.

The questionnaire was structured in such a way that actual instructional usage was followed by indication of the subject field of utilization. (See the Questionnaire in Appendix B.) Consequently, as regards GIDC, the subject areas that are mentioned most often would also be the subject areas that receive the most instructional usage. Such a conclusion, however, must be tenuous in the light of other factors that also affect usage. For example, the subject area of education was the largest area indicated with 53 persons, or 28.1% of those indicating instructional usage of the information they received, coming from that field. The area of education is supported by a large, well-developed, and apparently well-known data base, that of ERIC. (Indeed, a number of respondents recognize the GIDC system only in terms of ERIC. As one said, "I do not know what the GIDC is, but I have used ERIC", and another said, "The only service I have ever used is ERIC.") So the extensive usage in the area of education may simply reflect the existence of an adequate data base, rather than any particular characteristic of the field itself which might be unusually appropriate for computer searches.

Also the departments of education within the university system which the GIDC serves, generally have large enrollments and consequently more faculty. So the large number of users in a particular subject field might simply reflect an aspect of the user population, such as size, rather than an aspect of the subject field itself.

After education, biology was the second largest subject mentioned (by 33 or 12.2%), and chemistry was next (by 120 persons or 10.6%). Here again this usage could reflect somewhat the existence of adequate data bases. However, psychology, which was indicated by 18 users, or 9.5%, as their utilization area, and which was the fourth highest subject mentioned, was named under Question 8 by at least four users who expressed the desire for better coverage in the area of psychology.

Geology, social science, physics, agronomy, pharmacology, biochemistry, veterinary medicine, entomology, and microbiology were all mentioned between five and eight times each as being the subject of the information utilized, while a large variety of other fields, such as economics, foreign language, agriculture, and information sciences, were mentioned once or twice.

The questionnaire suggested four different types of courses in which the user might have used the information he received from GIDC: lecture, seminar, special project, and laboratory. The special project seemed to be most adaptable to the GIDC information, for it was checked by 52.6% of those who indicated instructional usage. After that came the lecture, with 40.9%, and then the seminar, with 35.1%. Laboratory usage was indicated by 24.4% of the users. The category of "other" covered such methods as workshops, individualized instruction, thesis research, discussion groups, graduate research, and research papers.

Most of the educational users (136 or 72.2%) indicated that, for their purposes, it was necessary for them to obtain full-text documents rather than to rely on titles and abstracts, although 81 did say that titles and abstracts were sufficient. In a few cases there seemed to be some confusion about the availability of abstracts and methods of securing them.

The last two questions, 7 and 8, dealt with suggested improvements in the GIDC service that might increase its instructional utilization. The original purpose of the questionnaire, to determine the amount and kind of usage for instructional purposes that the Center is receiving, must be kept in mind in looking at the responses to these queries. Particularly in relationship to suggestions for improvement, important consideration was not given to the possibility, advisability, practicality, or likelihood of implementation; the effort was merely to ascertain what changes the user felt would enhance the utilization of the system for instructional purposes.

Question 7, a short list of changes or improvements, was answered in the following numbers:

More descriptive abstracts	39.2% or 161 users
Browsing capability	29.5% or 121 users
Interactive system for query or profile formulation	27.3% or 112 users
Easier access to the system	22.4% or 93 users
Shorter waiting time for information delivery	22.4% or 92 users

Question 8 asked for suggestions from the reader for the improvement of service, and it is significant that a large measure of satisfaction with the service was indicated by a number of users. Favorable comments, such as "excellent service", "time saver", "most useful" and "very helpful", appeared on some of the questionnaires, and while there were a few negative comments too, the inclusion of suggestions for improvement should not necessarily be interpreted as unfavorable criticism.

The suggestions themselves fall into two general areas, those having to do with data bases and coverages of the service and those concerned with the functional details of the operation.

The most repeated suggestion involving data bases was for greater coverage. Twenty persons, or 4.8%, desired expansion of data bases, and the areas of the social sciences, the humanities, and psychology were each mentioned several times as being areas in which coverage should be expanded.

The matter of language was a concern for at least five respondents who requested that the language in which the article appears be specified in the entry and that the language of articles be specified in such a way that items can be chosen or rejected on the basis of particular languages. Another user asked that less from Russian journals be included.

Eight persons requested that more complete information, including addresses of authors, be given to facilitate ordering reprints.

At least five persons suggested more abstracts, although two suggested the use of summaries or annotations instead of abstracts. Another would like to access abstracts of dissertations.

Obtaining materials once they have been cited is a problem for some, and suggested solutions included adding directions for securing articles to the citation, some indication of the availability of the article in local libraries, or including the Library of Congress call number with the citation.

Books, as opposed to journal articles, are sometimes not included in certain data bases, and at least two users would like to have books, their contents, and their prices, as a part of the coverage. Another would like to see Books in Print available for computer searching.

One user would like some sample profiles with print-outs to use in instruction. Another wondered if permanent files of subjects frequently requested for retrospective searches might be a time and money saver. A two year record of on-going research in various fields would be an asset for still another user, and the great amounts of paper used, with the expenditure in time and trees, were a cause of at ast one complaint.

The suggestion most often made in the area of operational details of the GIDC was that the system should have more and better publicity to create a greater awareness of the system. At least 25 persons cited this need. Perhaps related to this suggestion is the fact that 10 respondents stated that they are unfamiliar with GIDC ("I don't know what you are talking about", "I never heard of GIDC."), although names of all recipients of the questionnaire were taken from the GIDC users list.

The second most frequently made suggestion (from 8 persons) was for more education or training for the user in order to give him some understanding of profile preparation, the strengths and limitations of computer searches, and his own responsibility in modifying and adjusting profiles to improve their effectiveness.

The mechanics of profile preparation was the object of a number of suggestions. Eight persons mentioned the need for more specificity in the profile. The ability to interact with the system and to do "trial and error" searches were mentioned as making for possible improvement in profile preparation. One professor would like some sample profiles set up to run just for teaching, and another would like to be able to run a quick, one-time profile for rapid turn-around. An annual request from the Center for profile changes and adjustments was also suggested.

At least 3 users complained of the excessive time required to file a profile and to begin receiving some search results. A suggestion for improvement in that area was to print out search results locally (i.e., on each campus) rather than at a central location as is presently done.

Five persons felt that easier and faster access to personnel that provide service for the Center would be an improvement. Simply getting an appointment with the person who assists in profile preparation was a problem for two users, and one user was frustrated because he felt that the person assisting him was not sufficiently knowledgable in his subject to give him the assistance he needed. Relating the service more closely to the library and to the manual searching there was another suggestion, and one graduate student felt that one method of improvement would be to permanently assign specific subject areas to specific librarians for profile preparation and for retrieval assistance as well.

Locating information after it has been cited created problems for some, and the circulation of an annual list of journals searched with an indication of their availability in the local library could be a partial solution.

Other suggestions that were mentioned included the desire for a consistent search schedule to produce profiles on a regular basis rather than erratically, the need of some method to request particular abstracts after the first bibliographic profile has been scanned for relevancy, and some way of reducing the amount of paper used.

In conclusion, an assessment of what was learned from the questionnaire in terms of its intent should be made. It seems clear that some educators are using the GIDC for some instructional purposes, primarily those tasks that utilize bibliographic information. While the suggestions that these educators made for improvements in the service of GIDC were numerous and varied in scope, their implementation would not seem to produce any appreciable increase in the instructional usage of the Center.

PART FOUR

DESIGN OF STITE DATA BASE MANAGEMENT SYSTEM

DESIGN OF STITE DATA BASE MANAGEMENT SYSTEM

L. J. Gallaher and Pranas Zunde

I. INTRODUCTION

The major functions of the STITE system were broken down into tasks and discussed in the second STITE Progress Report (3). Seventeen major tasks were identified and analyzed. All of these center around the use of the "internal" information, structured as modules and described in the third STITE Progress Report (4).

The main function of such modular "internal" information is to enhance systems applications, particularly in retrieval of "external" information. This information, stored in various scientific and technical information centers, consists normally of standard bibliographic data, keywords used to index documents, and in some cases, abstracts. Henceforth this information will simply be called "external data".

This report will describe programs which are being developed for STITE data base management, including both internally and externally retrieved information. In particular, the physical and logical structure of the files, together with a command language for manipulating the data structures in the files, will be discussed.

II. DATA BASE ELEMENTS

Data base elements are records of two kinds: modules of internal information and units of external information, such as bibliographic records.

The modules have been described in detail earlier. Each consists of textual material and additional information about the textual material.

This information associated with the text is arranged into fields, with the principal topic being designated as A field. B field contains a set of terms which are used to describe or to explicate the principal topic, or A field. C field indicates the source document from which the textual material is taken, F field specifies the form of the text (i.e., English language, chart, etc.), H field indicates type of information (i.e., definition, theorem, or problem), and D field refers to the level of difficulty of the textual material.

Modules can be ordered or linked to each other by various relationships defined on these fields, and structures can be built with the modules or their components. It is these structures or relationships that are of prime interest and that will be used to aid in selecting the bibliographic information.

External data would normally be supplied by an on-line link to various data bases, such as the Dialog (Lockheed) system or the Georgia Information Dissemination Center. However, for demonstration purposes, these external sources will be simulated by appropriate material that is in fact internal to the system since such an on-line link is not economically feasible within the scope of the financial resources of the STITE project.

External data will be processed by the STITE system in two steps, or levels. At the first level, external data will be retrieved by queries submitted to the external source in the standard format expected by that source. At the second level, information retrieved at the first level will be processed for the purpose of (1) achieving compatability with the records of the internal system, (2) extracting additional information, and (3) updating lists and file structures.

At the first level, the main task is to present an intelligent request to external sources using the internal information stored in the modules, and at the second level, the main task is to cast the records retrieved from external sources into a format approaching as closely as possible the format of complete modules.

III. GENERAL DESCRIPTION OF THE STITE INTERFACE SYSTEM

Figures 1 through 5 are diagrams displaying main features of the overall STITE system. These diagrams represent the system from the point of view of the user and emphasize the interaction with the user.

Information about the system, its capabilities, and the type and extent of its information stores is incorporated into the system and will be supplied to the user by the system in a dialogue form.

Explanations to the user about the system are of two kinds. The first kind is a general outline of what the system is supposed to be able to do for the user and how to go about using it. This is the "user's manual" and will give an explanation of the details for using the system. It will be structured hierarchially so that the user may skip through it to the points of interest to him and is similar to the standard introductory material needed for any on-line system.

In addition to the general outline on the use of the system, there will be a region of more specific information about the subject matter contained in the modules. Here the user will be interrogated about his specific interests as regards subject matter and task and an effort will be made to inform him of the content of the modules and how he can or cannot be helped by the system. The emphasis here is on subject matter and on what subjects are covered by the modules and what are not.

To carry out certain tasks, especially the open-ended ones, a command language will be provided to the user after he arrives at that task. This is

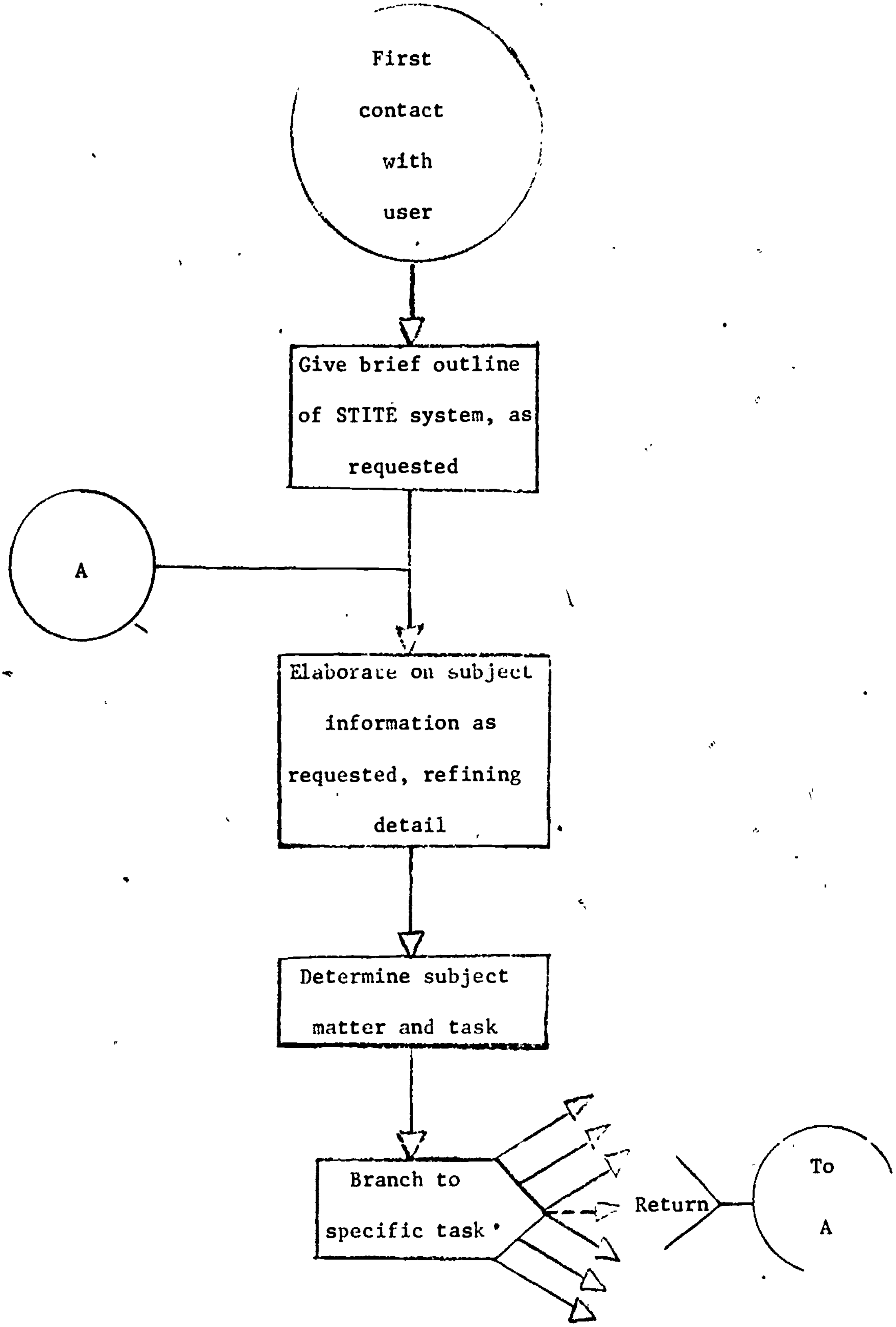

First contact with user
Give brief outline of STITE system, as requested
A
Elaborate on subject information as requested, refining detail
Determine subject matter and task
Branch to specific task
Return
To A

Fig.1

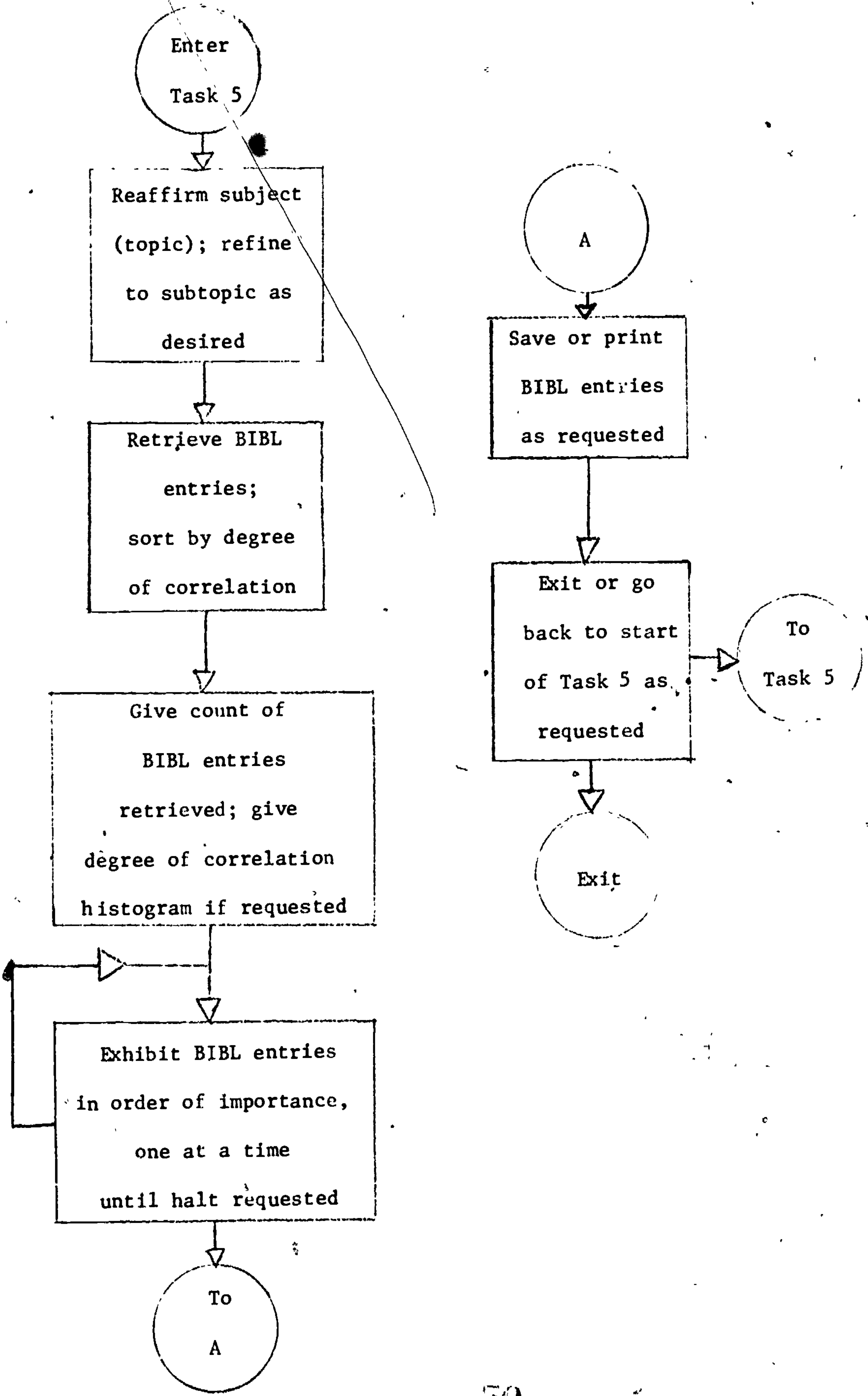

Enter
Task 5
Reaffirm subject (topic); refine to subtopic as desired
Retrieve BIBL entries; sort by degree of correlation
Give count of BIBL entries retrieved; give degree of correlation histogram if requested
Exhibit BIBL entries in order of importance, one at a time until halt requested
To
A
A
Save or print BIBL entries as requested
Exit or go back to start of Task 5 as requested
To
Task 5
Exit

Fig.2

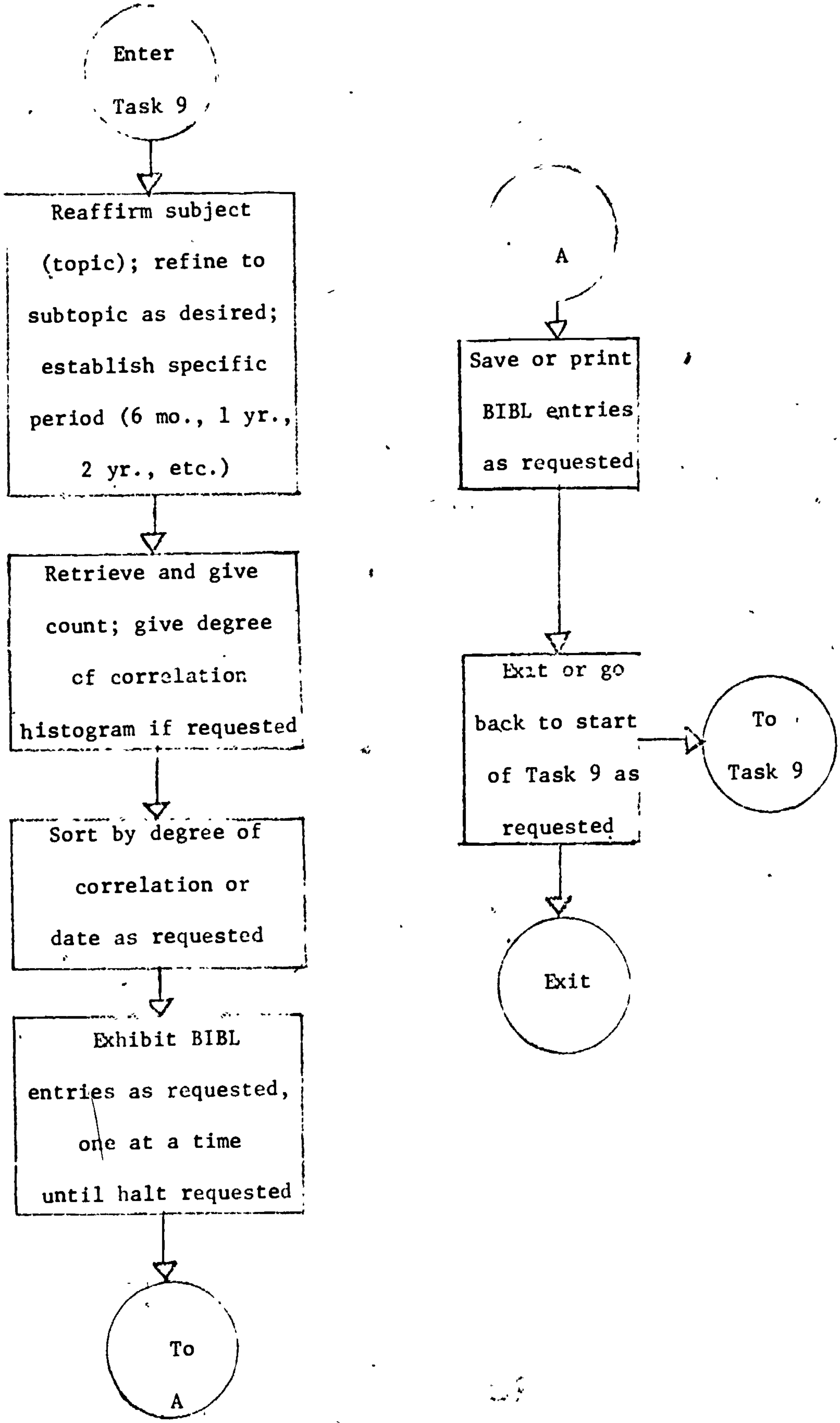

Enter Task 9
Reaffirm subject (topic); refine to subtopic as desired; establish specific period (6 mo., 1 yr., 2 yr., etc.)
Retrieve and give count; give degree of correlation histogram if requested
Sort by degree of correlation or date as requested
Exhibit BIBL entries as requested, one at a time until halt requested
To A
A
Save or print BIBL entries as requested
Exit or go back to start of Task 9 as requested
To Task 9
Exit

Fig. 3

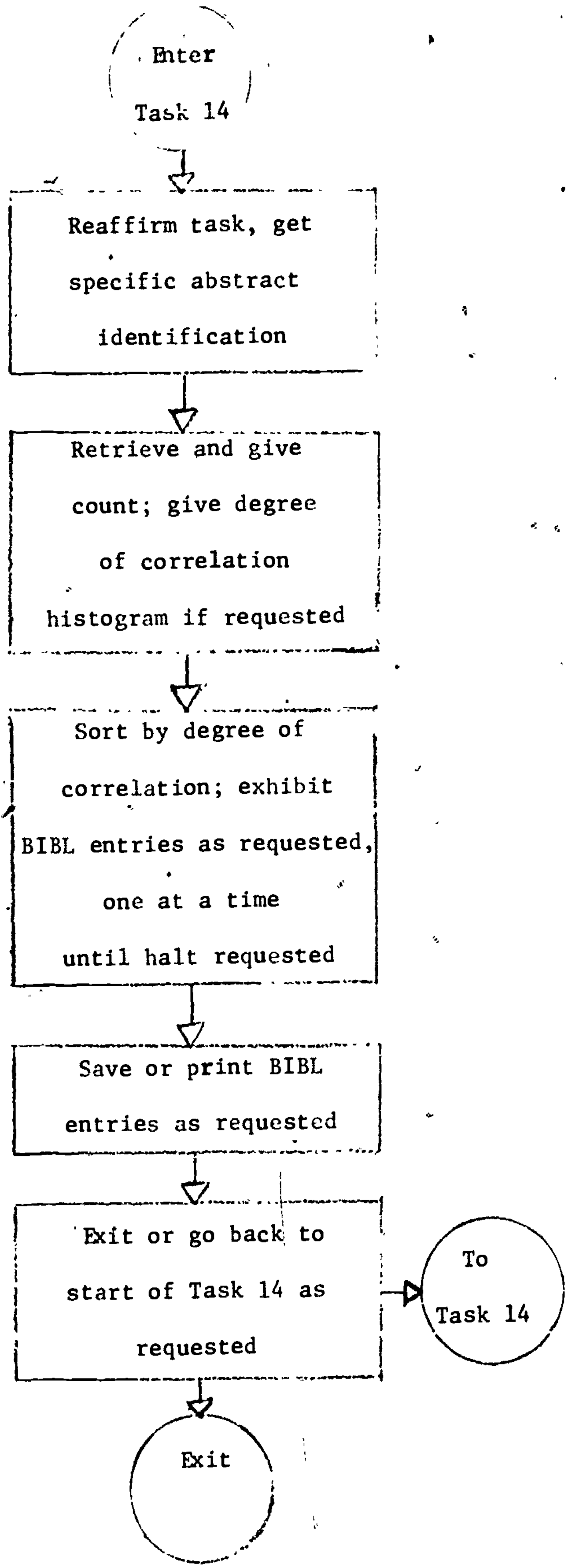

Enter Task 14
Reaffirm task, get specific abstract identification
Retrieve and give count; give degree of correlation histogram if requested
Sort by degree of correlation; exhibit BIBL entries as requested, one at a time until halt requested
Save or print BIBL entries as requested
Exit or go back to start of Task 14 as requested
To Task 14
Exit

Fig.4

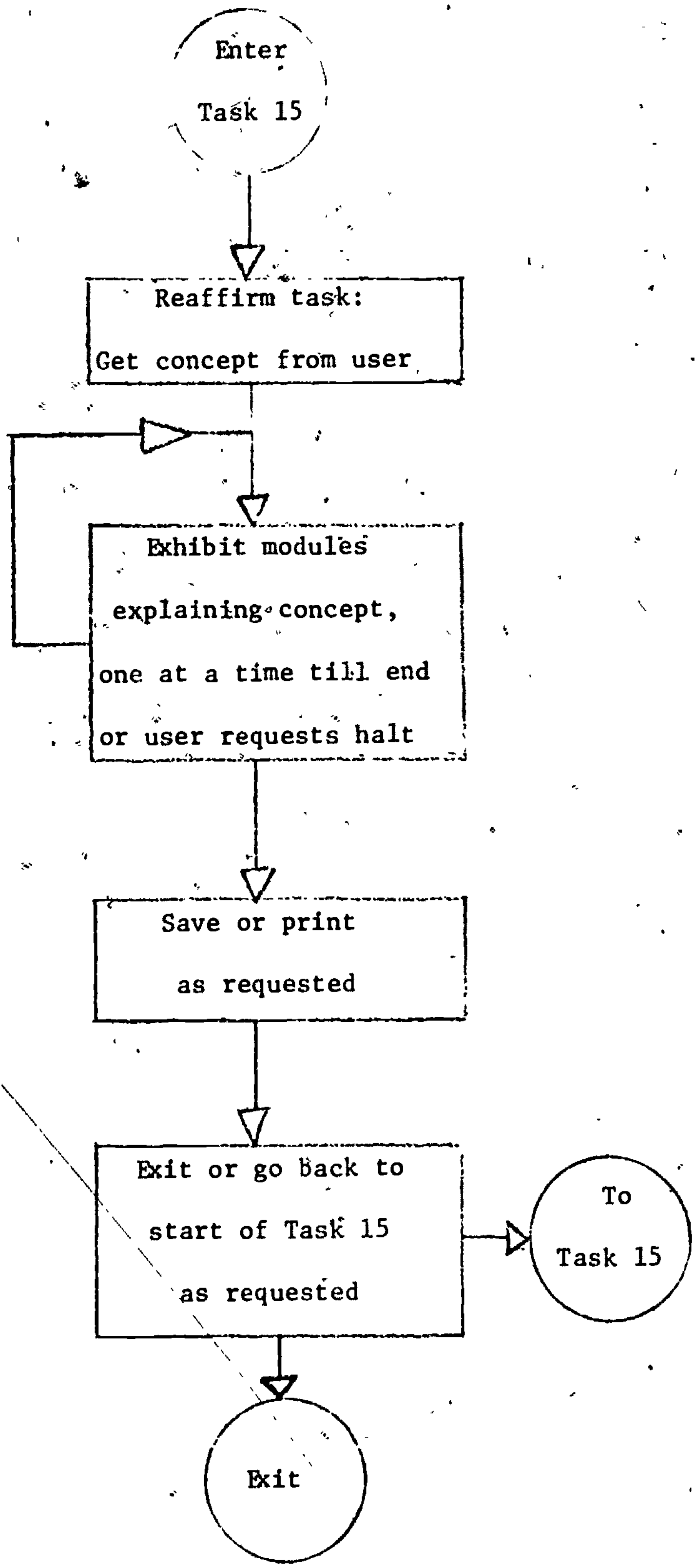

Enter Task 15
Reaffirm task: Get concept from user
Exhibit modules explaining concept, one at a time till end or user requests halt
Save or print as requested
Exit or go back to start of Task 15 as requested
To Task 15
Exit

Fig.5

not a conversational language but a set of commands that allows him to manipulate, build, and rearrange file structures to suit his needs. This command language itself is quite simple and will be interpreted by the system.

At this stage, the following tasks have been considered from the list of potential tasks described previously in a STITE Progress Report (3).

Task No. 5 - Compilation of bibliographies on selected topics.

Task No. 9 - Compilation of references for a state-of-the-art review or of abstracts covering a specified time period.

Task No. 14- Retrieval of all abstracts related to a specific abstract.

Task No. 15- Presentation of concept definitions and explications.

IV. THE PHYSICAL AND LOGICAL STRUCTURE OF THE FILES

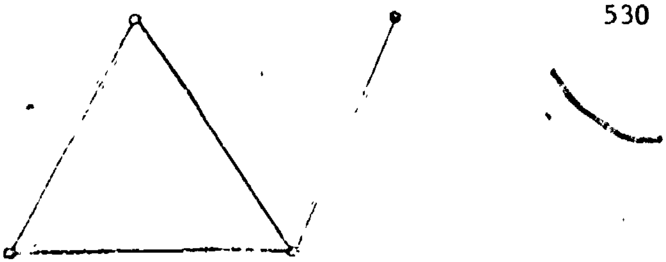
Previous reports have outlined in a general way the organization of the automated STITE system. Details of the structure of the information files, which will make possible the kinds of action desired, will now be given.

There are two distinct kinds of files: records of internal information, or modules, and bibliographic data records (BDR) embodying external information.

A bibliographic data record indicates the author and title of the document title, date, and page number for a journal article, report number and sponsoring agency for a technical report, eventually an abstract of the document, a list of keywords which have been used to index the document, and the index or abstract journal, with its date and abstract number in which the document appears.

The information content of each module has been previously described in another STITE report (4). Figure 6 below illustrates typical module content.

530



A	B	C WIL/1/29
simple graph	vertex-set	
	edge-set	F 2
	pair	
	edge	H 7
	ex	
	se	
	family	D 1

629

Given any graph G, an edge-sequence in G is finite sequence of edges of the form

$$\{v_0, v_1\}, \{v_1, v_2\}, \ldots, \{v_{m-1}, v_m\}$$

(also denoted by $v_0 \, v_1 \, \ldots \, v_m$). It is clear that an edge-sequence has the property that any two consecutive edges are either adjacent or identical; however, an arbitrary sequence of edges of G which has this property is not necessarily an edge-sequence.

A	B	C WIL/1/128
edge-sequence	finite sequence	
	edge	F 1
	form	
	consecutive edges	H 1
	adjacent edges	
	identical edges	D 1
	arbitrary sequence	

Modules 530 and 629

Fig.6

The fundamental unit of a file is the record which, if not specified, refers to either a module or a BDR. The records of both the module and BDR files are organized in a similar manner and so will be discussed together.

A record is considered to have a name and an association list. The record name will be an alphanumeric string consisting of a number prefixed by BDR if it is a bibliographic data record or by MOD if it is a module. The association list will consist of a collection of pointers designating the elements (properties) or fields of the records.

Each record consists of a collection of fields. Each field has a name, value, and association list. The fields are the sub-elements of the records and differ from the records in having values in addition to names. The name and value of a field are joined in a single alphabetic string with the value set off by curly brackets, "{}". Field names consist of a one, two, or three letter string. For example, the name value for the A field of module 629 of Figure 6 would be A {edge-sequence}, the B field name value would be B {finite sequence}, B {edge}, B {form}, etc.

Most of the records also have free text entries, and these are referred to as the free text fields and are treated slightly differently.

Each field value (except the text) also has an association list. Association lists of the field values can have such things as the list of all records containing that particular field value (the inversion) or any other information deemed relevant or useful. Initially most of these lists are null, but the capability of an association list for a field will always be present except for the free text field. The free text fields do not have association lists.

The concept of field is quite general so that it is possible to assign new fields to records as needed. Also, in discussing the structure of the files, it will be seen that the distinction between file, record, and field is rather artificial and that all "logical" units are treated in a unified manner.

Physically the files are divided into two kinds: alphabetic information (called α-files) and pointer information. The alphabetic information is stored as alpha-numeric strings with distinct units separated by a special character (low value). The pointer information is stored as a collection of binary (machine language) integers.

Alphabetic files contain all the names, values, and free text information. The pointer information shows the form of the association lists and indicates such things as the structure of a record or file.

The pointer information is of two types also: hash references and association lists. These are not separated since the structure of the two is very similar, as well as the frequency and mode of reference.

Figure 7 gives a schematic representation of the file structure and linkages for a part of record Module 629. Details on the hash file and the structure of the association lists will be given later in this chapter.

Figure 8 represents another view of the file structure. Here the files are considered in terms of ownership and are to be made up of permanent and temporary working areas. From this point of view, the file owners are the system (the system manager) and the users. The system manager can be considered as a special user who owns the STITE system files. In this picture the BDR files are external to the system and represent the information available from the outside and are "owned" by their supplier.

The users will have both temporary and permanent files. They will also have the capacity of moving one file into the other but with this

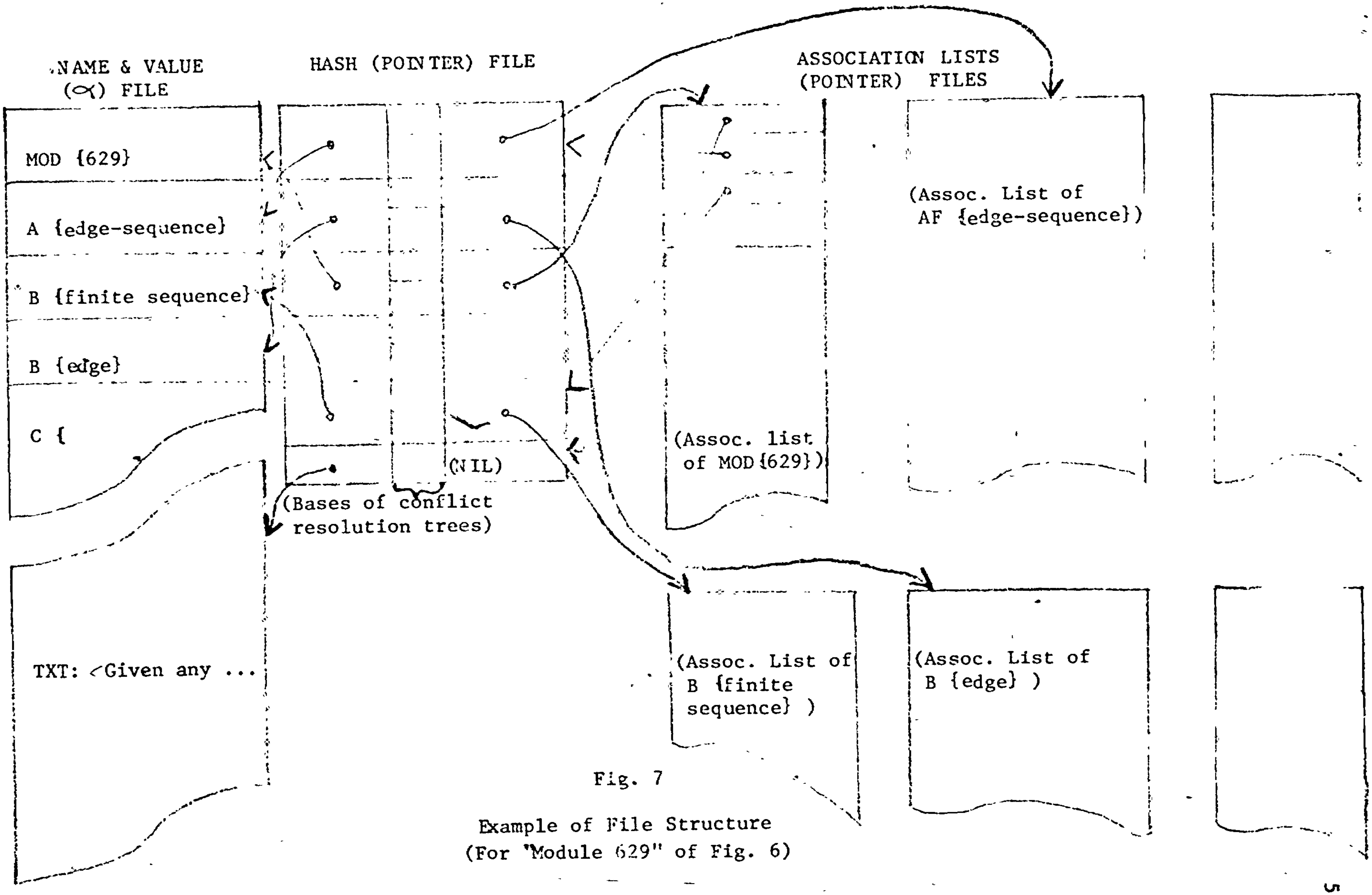


Fig. 7

Example of File Structure
(For "Module 629" of Fig. 6)

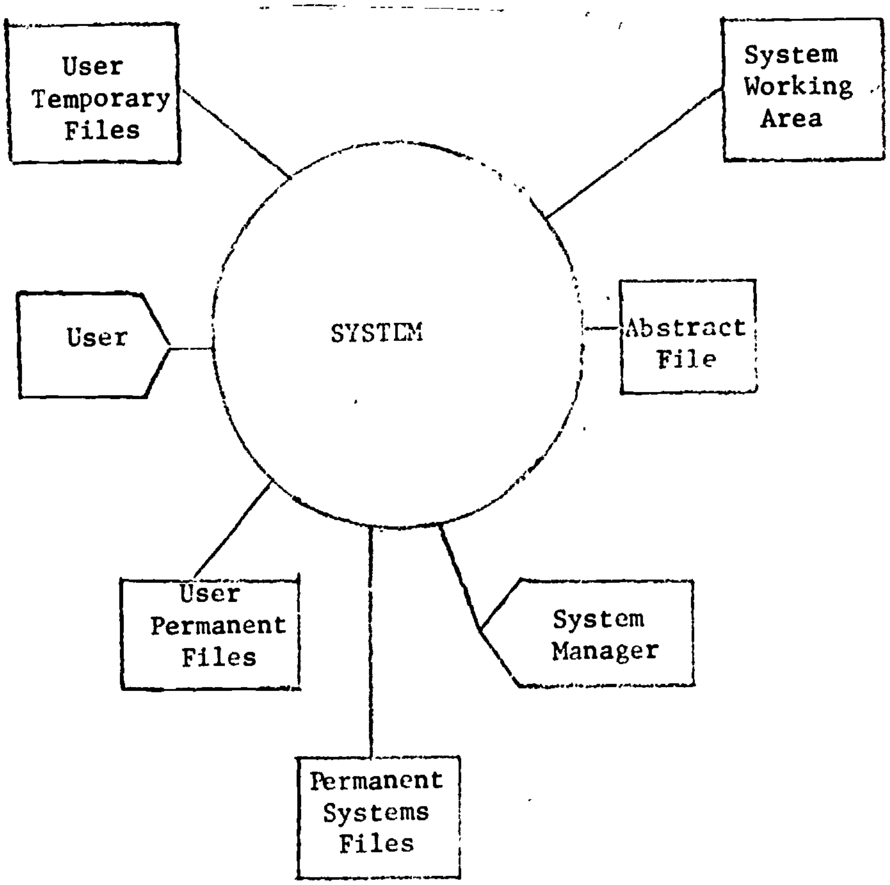

User Temporary Files
System Working Area
User
SYSTEM
Abstract File
User Permanent Files
System Manager
Permanent Systems Files

Fig.8

restriction: A user may not move a file into a file not owned himself. Thus, only the system manager can modify the permanent system files.

The main characteristic of the temporary files is that they disappear when the user disconnects, while the information in the permanent files is kept and is available again at the next use. This ability to create structures in temporary files and then to move to permanent areas is an important capability and one that will allow the effectiveness of the system or of a user to grow with time.

The distinction between the term "physical" structure and "logical" structure in a file system such as this is important and needs clarification. This is particularly true when one recognizes that the files are handled through an operating system which in turn makes its own distinctions. The term "physical file" is used to mean a named file, viewed as a distinct entity by the operating system. The actual physical device assigned by the operating system is of no concern. However, in all cases, it is expected that the physical device is a disk and that the physical records are randomly accessible.

A "logical file" means those elements (i.e., free text, names, values, and pointers showing structure) that logically belong together whether they reside in the same physical file or not. From this point of view there are two kinds of logical files in the system, BDR files and module files, with further logical distinction associated with the owner and whether the file is permanent or temporary.

In this way it can be seen that a single logical file (or record) may be distributed over several physical files while at the same time a single physical file can contain components of several logical files. From the overall view of the operation of the STITE system, the logical file is the important concept. However, in dealing with the details of how the system is maintaining the information, the form of the physical files is important.

V. THE COMMAND LANGUAGE

For carrying out certain tasks and for performing other desired operations, a command language for manipulating the records and creating structures has been implemented.

Looking at the records or structures being manipulated, one notes that they are basically names and sets of names. Associated with each name is a list called its association list. An association list is just a set of names (since the order is not relevant and no duplicates will be allowed). Some of the association lists are empty. Set theoretical operations are defined and implemented on these association lists, i.e., on these sets of names.

The commands are in a very simple form and restricted to unary and binary operations. One of the operands is designated by a P or by pointing register; the other, if there is one, is named in the command. (This is reminiscent of a single address machine language with an "accumulator" register, except that the "accumulator" register does not hold the item to be operated on but only points to it.) The item pointed to by the P or pointing register will sometimes be referred to as just P. Thus, the command, point X, sets P to X, that is, sets the P or pointer register pointing to X. The command attach Y adds Y to the P list, that is, adds the name Y to the association list of the name pointed to by the P or pointer register.

This command language is outlined briefly in Table A.

Some examples of the use of these commands to perform some elementary operations follows:

1) Create Q so that it is the union of X and Y. That is, the association list of Q is the union of the association lists of X and Y.

```
create Q
add list X
add list Y
```

2) Remove all items from the association list of R, i.e., make R the null list:

```
point R
delete list R
```

or

```
create Z
point R
delete list not Z
```

3) Display those items on the association list of X but not on the association list of Y:

```
create TEMP
add list X
delete list Y
display list
```

4) See if X is on the union of Y and Z:

```
create T2
add list Y
add list Z
is member X
```

Following each command is a response by the system to the user giving an indication that the command has been carried out or cannot be obeyed. Examples of commands that cannot or will not be carried out are as follows:

1)	create X	Not allowed if the name X already exists as a recognizable name
2)	point X	Not allowed if the Y does not already exist as a recognizable name
3)	remove X attach X add list X etc.	Not allowed if the name X does not already exist as a recognizable name, or if P register has not yet been set
4)	display display list count	Not allowed if pointer or P register has not yet been set to point to a recognizable name

While this ia a very primitive type command structure for manipulation of the association lists (sets), it allows the implementation of the desired processes. Such a simple scheme was chosen for ease and speed of implementation without the need for a study of language implementation and compiler writing techniques which are not directly among the goals of this project.

VI. THE HASHING PROCEDURE

Earlier in this section it was mentioned that the alphabetic information was entered and linked by a hashing algorithm (Figure 7). This hash algorithm is in three stages.

First, the alphabetic string is mapped into an integer, I, by a suitable procedure that tries to give a uniform distribution over some interval for the character strings to be encountered. There are many ways of doing this mapping, and nearly any of the traditional methods would be satisfactory. (The actual scheme used here was to do exclusive OR's of the 7 bit (ASCII) patterns of the characters, taken pairwise. The resulting 14 bit pattern is taken as a binary integer on the interval 0 to $2^{14}-1$).

Second, this integer, I, is then reduced modulo the hashing interval, M, (by a remainder divide) to give the initial hash position. If this initial hash position is empty, the item being hashed is placed there. Otherwise proceed to the third step.

TABLE A

COMMANDS FOR BUILDING AND MANIPULATING ASSOCIATION LISTS

CREATE X	Create the name "X", associate the null list, set P register pointing to X
POINT X	Set P register pointing to X
ATTACH X	Put X on the association list of whatever is pointed by P register
REMOVE X	Take X off the association of whatever is pointed by P
DISPLAY	Display name pointed by P
DISPLAY LIST	Display the association list of whatever is pointed by P
ADD LIST X	Add the items on the X association list to the items on the list pointed by P, eliminating duplicates (the union or OR operation)
DELETE LIST NOT X	Delete from the list pointed by P all items not on the X list (the intersection or AND operation)
DELETE LIST X	Delete from the P list all items on the X list (the AND NOT operation)
COUNT	Counts the number of items on the pointed or P list, 0 if null
IS MEMBER X	Responds <u>yes</u> if X is on the pointed (P) list, <u>no</u> otherwise
RELATION LIST X	Gives set relation between the pointed list and the association list of X. Response: "Identical" if X & P lists identical "Superset" if X is a proper superset of P "Subset" if X is a proper subset of P (Count) of intersection of X & P "Disjoint" if X & P lists disjoint
MAKE LIST NULL	Sets the P list to null
ON SECONDARY PUT X	X is placed on the lists of the items on the list of P
UNION SECONDARY	The list of P is replaced by a list that is the union of lists of the items on the original list of P

TABLE A
(continued)

INTERSECT SECONDARY	The list of P is replaced by a list that is the intersection of the lists of the items on the original list of P.
ABITERATE	The list of P must consist of A field name values (i.e., of the form A{ ... }). The list of P is replaced as follows: 1) find the union of all the modules for the A fields, 2) find the all B fields of all these modules, 3) convert all these B fields to A fields (with the name values). The final P list is the set of all the A field name-values obtained in step 3.
BAITERATE	Same as ABITERATE except that the A and B fields are interchanged.
BBITERATE	Same as ABITERATE except that one starts with B fields (and ends with B fields).
ABCONNECT X	Performs ABITERATE until a match on X is obtained; after each iteration the number of items in the P list is given and the option of continuing or stopping is given.
BACONNECT X	Same as ABCONNECT except with A and B fields interchanged.

Third, hashing conflicts are resolved by building a binary tree out of the initial hash position. The choice of going to the right or left branch of the tree is determined by the odd or evenness of $\lfloor I/2^j \rfloor$, where j is the node level of the tree. The item is put in the first empty position encountered.

Figure 9 gives a pictorial representation of this hashing scheme. This algorithm can be expected to give uniform distribution over M and relatively well balanced trees if M is prime, small with respect to Imax (the maximum of I) and the I's are distributed uniformly over the interval 0 to Imax. In this application Imax = $2^M - 1$ and M is a program parameter normally set around 2^{11}. The actual computer program is further complicated by the paging of the links which makes it desirable to make M the product of a number of pages times the number of links used per page. Thus, instead of M being prime, it is the product of two primes, one of which is 41, the number of links used per page.

The final position in the hash table or tree of an alphabetic string will be referred to as its internal name, its external name being the alphabetic string itself. The programs then deal with this internal name wherever possible, using the external name only when communicating with the user. Note that once assigned, the internal name is unique.

VII. ASSOCIATION LIST STRUCTURE

The association lists are also trees, being in this case ternary trees. (Ternary trees were chosen for convenience; this makes the linkage sizes for the two groups of linkages, the hash tables and trees, and the association lists compatible. Thus, it is convenient to mix these two in the same physical file and buffer areas.)

THE HASHING TREES FOR ALPHABETIC STRINGS

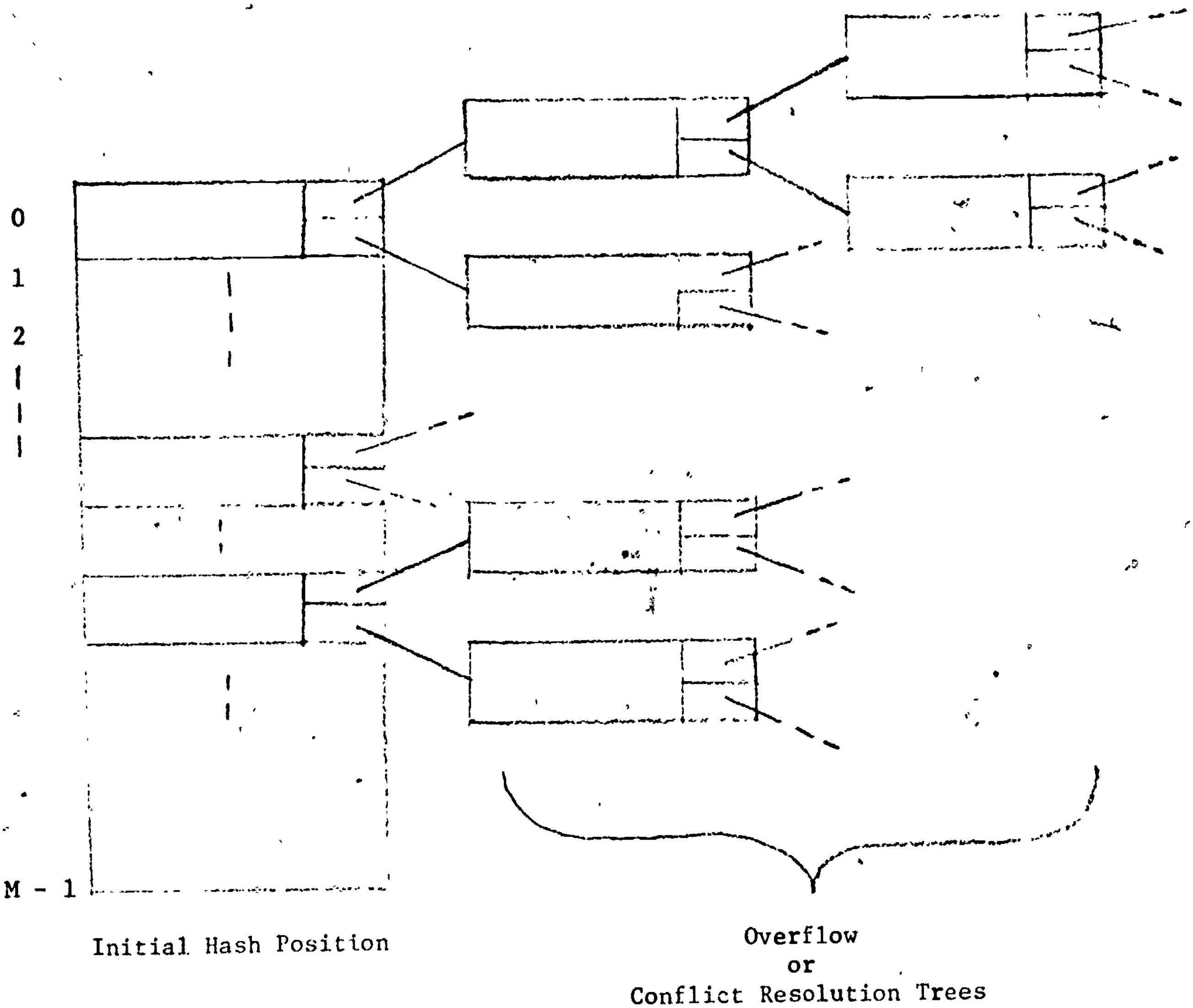


Fig. 9

The algorithm for placing the entries on the tree is "hash" like: The name of the object being placed (in this case always the internal name) is mapped onto the interval 0,1,2 at each node. This mapping is pseudo random in the sense of giving 0,1, and 2 with equal frequency but is unique and reproducible for each internal name. Many schemes for doing this are satisfactory, and the one used here was chosen for programming convenience. It is best described by giving the Algol-like program for performing the choice at each step:

```
begin
if t≠0 then t:= t ÷ 3;
else begin
          j:= j + 1;
          t:= (j ⊗ line) V page
          end
k:= t mod 3
end
```

Here t and j are initialized to zero (at the base of the tree). The operator V is the bit wise exclusive OR operation, and mod is the remainder divide operation. Page and line are the (unique) page and line numbers of the internal name being entered on the list. k gives the branch on the tree at each mode, i.e. k = 0,1,or 2.

A few words about these "hashed" trees used for the lists as compared to sorted trees. First, one notes that the ternary trees are obviously faster to access than the corresponding binary trees. That is, the average number of looks per item sought is smaller by the ratio log 2/log 3. Second, one notes that hashed trees are on the average much better balanced than the sorted trees. In fact, hashed binary trees are on the average about as well balanced as (sorted) AVL trees, yet require no balancing.

Finally, sorted trees are normally binary; ternary sorted trees being unduly complex and less efficient.

Thus, one concludes that for single item look up, the hashed trees are superior to sorted trees and even to balanced sorted trees. Of course, there is a disadvantage in not having the information sorted, and retrieving large numbers of consecutive items from a hashed tree is not easy. Furthermore, the merging of two trees cannot make use of the merge algorithms for sorted trees. However, it is not anticipated that either of these two kinds of operations (interval retrieval or merging) will be particularly significant in the application of this system.

VIII. CHARACTERISTICS OF LINKAGE-PAGING SYSTEM

The computer programs for processing the records are written in the C language for the PDP 11 computer. C is an Algol-like language particularly well adapted to the PDP 11 computer and operates under the Unix operating system, a time-sharing system for the PDP 11.

The PDP 11 configuration in use has 104 K bytes of core and about 200 M bytes of disk storage, although, of course, not all of this is available to a single user. The maximum core allowed a single user by the Unix system is about 64 K bytes. It is anticipated that this project will use about 20 M bytes of disk storage.

Every effort is being made in the programming to use modular and hierarchical techniques. The C language is well suited to these methods.

Because of the limited core the first requirement of the programs is for a (data) overlay or paging system, and this is incorporated at the lowest level. Since it is the linkages that are expected to be referenced most frequently, these were incorporated into the paging system. The alphabetic information will be needed only for communication with the user and referenced as needed and is not involved in the "automatic" paging.

Many of the characteristics of the link-paging system are determined by the machine being used (PDP 11) and the operating system (Unix). The fixed page size of 512 bytes was chosen to fit the hardware/software. A link consists of four addresses, each address being 3 bytes, 2 bytes to indicate the page number and one for the line (or link) within the page. This allows the spanning of 64 K pages of linkage data, the maximum allowed to a user (by Unix). This also allows 42 lines (or links) per page with 8 bytes left over for some housekeeping chores.

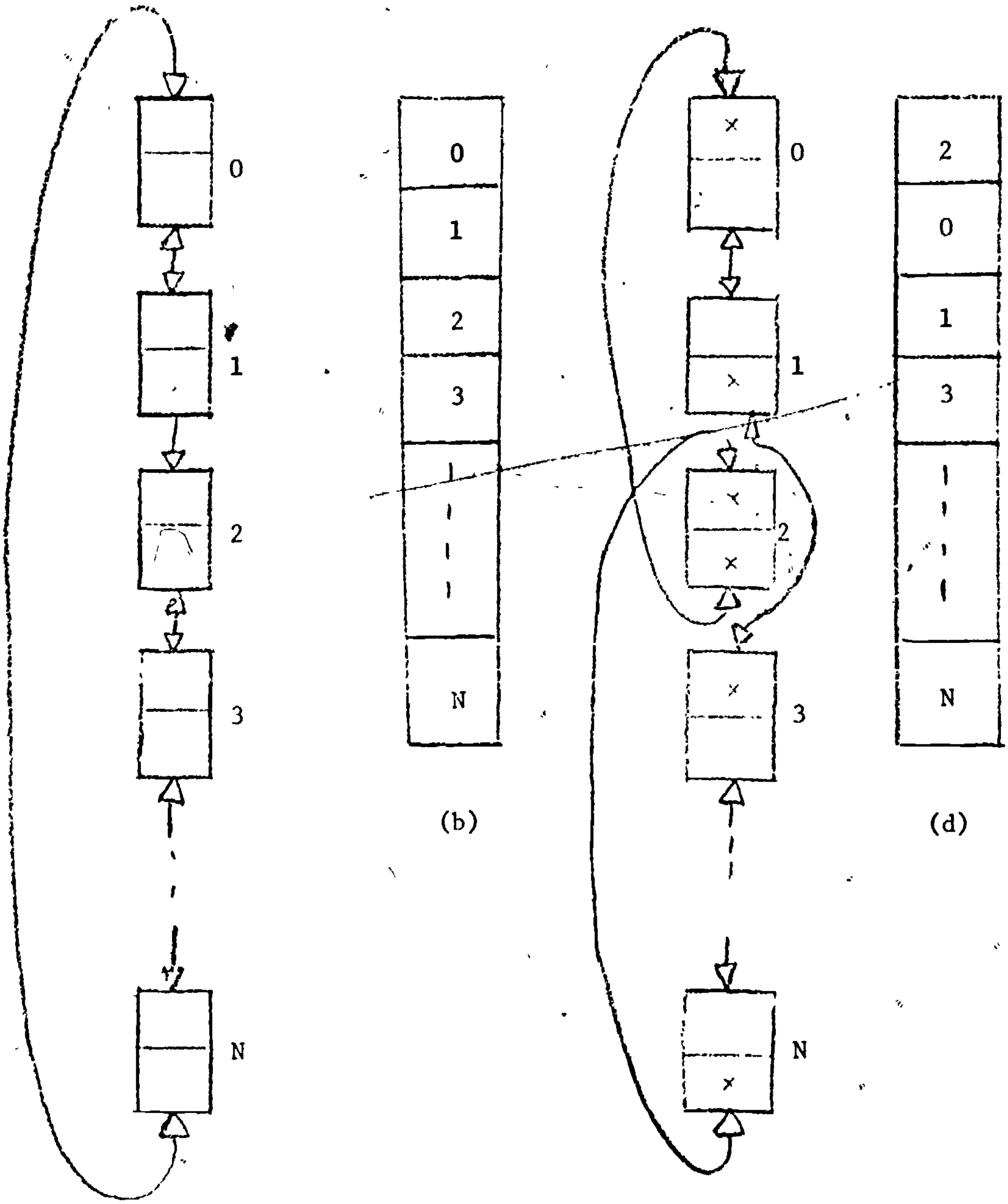
A system of N buffers is set up, with each buffer holding one page. N is a program variable and is expected to be between 32 and 64, consistent with other core requirements. The linkage pages are moved in and out of the buffer areas as needed with a directory kept of which are in core.

In addition, the directory keeps a record of the manner in which the pages are referenced. Those pages referenced least recently and/or least frequently are first to be moved out when a new page (a page not in core) is required to be moved in. Note this is neither a first-in-first-out nor a random paging system but a "least recent and least frequently referenced, first out" system.

The concept here is that during a given time interval in the running of the program, certain pages will be referenced with high frequency and others less often. The particular set of pages so referenced will change with time or with the demands of the program. The paging system is expected to adapt to these changes and to adjust in a statistical manner to the changing page requirements. It is also expected that the linkages will cluster so that there is a high probability that links will point to others on the same page. While no systematic efforts to produce clustering are made, this particular application and the manner in which the links are initially assigned will encourage clustering. However, it is recognized that if extensive changing and reassignment of the linkages occurs, the clustering or correlation of the links within a page will disappear and become random or incoherent. This could be corrected by a more sophisticated garbage collection and link assignment algorithms, but such techniques were not undertaken at this time.

The paging algorithm works as follows: There are N buffers, labelled 0 through N-1, each holding one page. Associated with each buffer is a pair of pointers designating a predecessor and a successor for that buffer plus one more pointer pair, not associated with any buffer, used to designate the first and the last buffer. The N+1 predecessor-successor pointer pairs are labelled 0 through N, and all but the last have an associated buffer. Initially, they are linked as in Figure 10a; 1 is the predecessor of 2, 2 is the predecessor of 3,..., 1 is the successor of 0, 2 is the successor of 1,...,. This is usually referred to as a doubly linked circular list. The arrowed lines show the predecessor-successor relation. Note that N is the predecessor of 0 and successor of N - 1 in Fig. 10a. Fig. 10b is another pictorial representation showing the buffers as a simple ordered set with 0 at the top and N - 1 at the tail of the sequence.

Now whenever a buffer is referenced, the pointers are changed so that the referenced buffer becomes the successor of N, the pointers being reconnected to show this shift. Thus, if buffer 2 is referenced, the



(a) (b) (c) (d)

Fig. 10

pointers will be changed to look like Fig. 10c. Considering the sucessor of N to be the head of the list and the predecessor of N to be the tail, the reconnection to Fig. 10c has promoted 2 to the head of the list while demoting by one position all those that were ahead of 2, without changing the position of the others. Fig. 10d is the other pictorial representation again showing the buffers as an ordered set with 2 at the top. One says that buffer 2 has been "popped to the top" of the list.

Note that nothing really "moved"; only three pointer pairs were broken and reconnected (three memory swappings or nine fetches and nine stores to reconnect the points). The pointers changed are indicated with an x in Fig. 10c. The algorithm in an Algol-like notation is as follows:

```
sn      := s[n];
pj      := p[j];
sj      := s[j];
s[j]    := sn;
s[n]    := j;
s[pj]   := sj;
p[j]    := n;
p[sn]   := j;
p[sj]   := pj;
```

Here sn, pj, and sj are temporary storage, s[j] and p[j] is the successor and predecessor the the j^{th} buffer, n is the last or N^{th} pointer pair, and j is the index of the buffer being promoted.

Each time a buffer reference is made, the referenced buffer is popped to the top of the list in the manner just illustrated.

This is the update procedure that keeps a record of which buffer or page was referenced last, which second last, which third last, and so on. The least recently referenced buffer will be the predecessor of N; its name or location is in the predecessor position of the N^{th} pointer pair and is easy to find.

Now the swapping algorithm is just that the page in the least recently referenced buffer is swapped out when a new page, a page not now in core, is needed. The old page is rolled out (if it has been written on), the new page is given this buffer location, and the pointers are reconnected so that the buffer of the newly brought in page is at the head of the list.

One final variation on this theme is added. Instead of having a buffer pop to the top on every reference, it may be moved up only sometimes. A pop to the top may take place on only every other, or every third reference. This will result in a probability distribution for the buffer positions in the swap out list. Those referenced with higher frequency have a higher probability of being near the top and those referenced less frequently will probably be near the bottom. Swap out takes place from the bottom of the list so that those pages referenced least frequently and/or least recently have a higher chance of being rolled out when a new page is needed. (It is assumed here that large numbers of page references are more or less random or at least have no systematic periodicities that might discriminate against certain frequently referenced pages being moved to the top of the list).

Whether pop to the top takes place every third or every second or on every page reference is a program parameter and can be set as desired. What is optimum in this application would need to be determined by extensive experiments. Such experiments are far afield from the real goals of this project and will not likely be pursued at this time. However, preliminary analysis suggests the following: If the usual situation is that page references are highly clustered (that is, if many successive references to the same page are likely), then pop to the top every third or fourth or even tenth reference would be more efficient than promotion on every reference.

If, at the other extreme, the correlation is very low or nearly random and pages are seldom referenced more than once or twice during their lifetime in core (i.e., before they are swapped out again), the pop to the top on every reference would be more efficient. One expects the situation here to be somewhere in the middle of these extremes so that promotion every second or third reference will be tried to start with, and performance of the overall system is not expected to be very sensitive to this parameter.

IX. STRUCTURE OF THE PROGRAMS

The C language (1) is a procedure or subroutine oriented language and well suited to modular and hierarchial programming techniques. Here will be given some of the structural characteristics of the programs.

The basic building unit is the procedure. A procedure is a more or less closed set of code that is called or invoked as a unit. It can have within it calls on other procedures (including itself). Procedures communicate with each other both through parameters and global variables. (While it is considered that procedure communication through global variables is "dangerous", i.e., an error-prone method of doing things, it would be most difficult and awkward to avoid in the C language. The best that can be done is to keep global variables to a minimum.)

The procedures are grouped into sets (called blocks) that perform particular tasks. Sets of these blocks are then built into programs that perform higher order tasks. Sets of programs can also be grouped together to perform yet higher order tasks. The program set is the top of the hierarchy in the C-Unix system on the PDP 11 (2).

Figure 11 is a block diagram for the module processing program that loads the modules and casts them into their linked list structure form. Each block represents a collection of procedures used to perform the indicated tasks. Global variables are global only within a block and may not be referenced outside of that block. Procedures within a block may reference each other or those in a lower block. Thus, for example, the set manipulation procedures reference each other and those of link-paging and α-file management, but link-paging and α-file management never reference each other nor the global variables associated with each others blocks. This hierarchy structure makes debugging and trouble shooting easier since the blocks can be checked out in order from bottom to top.

(Note here that "proving correctness" of our programs is not being done in any sense, but rather only methods which produce fewer errors and make those errors that do occur easier to find are being adopted.)

Even the procedures within each block have a hierarchy structure. Fig. 12 is a lattice diagram for the link-paging system showing which procedures reference which within this block. Note that a procedure only references those below it in the diagram to which it is connected by a line.

Figures 13 through 24 give the flow diagrams for the individual procedures of the link-paging system. These are just the procedures indicated as "The Link-Paging Management Procedures" in the lower left box of the block diagram, Fig. 11, and in the lattice diagram, Fig. 12.

Figures 25, 36, and 43 give the lattice diagrams for the α-file management procedures, the hashing procedures, and the set manipulation procedures respectively. Figures 26 through 35 give the flow diagrams for α-file management, 37 through 42 give the flow diagrams for hashing, and 44 through 60 are the flow diagrams for set manipulation. Fig. 61 gives the lattice diagram for the module processing procedures while Figures 62 through 69 give the flow diagrams for each of these procedures.

Most of the blocks in Fig. 11 will appear repeatedly in the other program block diagrams since they are the basic units needed in all link, character and set manipulations of the programs. This is the reason for giving so much detail on these blocks here. For example, in Fig. 70 is given the block diagram for the interactive (set manipulation) program. One sees that it uses most of the same procedures used for module processing. Fig. 71 is the lattice diagram for the interactive procedures, and Figures 72 through 87 give the flow diagrams for each of the individual procedures.

THE MODULE PROCESSING PROGRAM

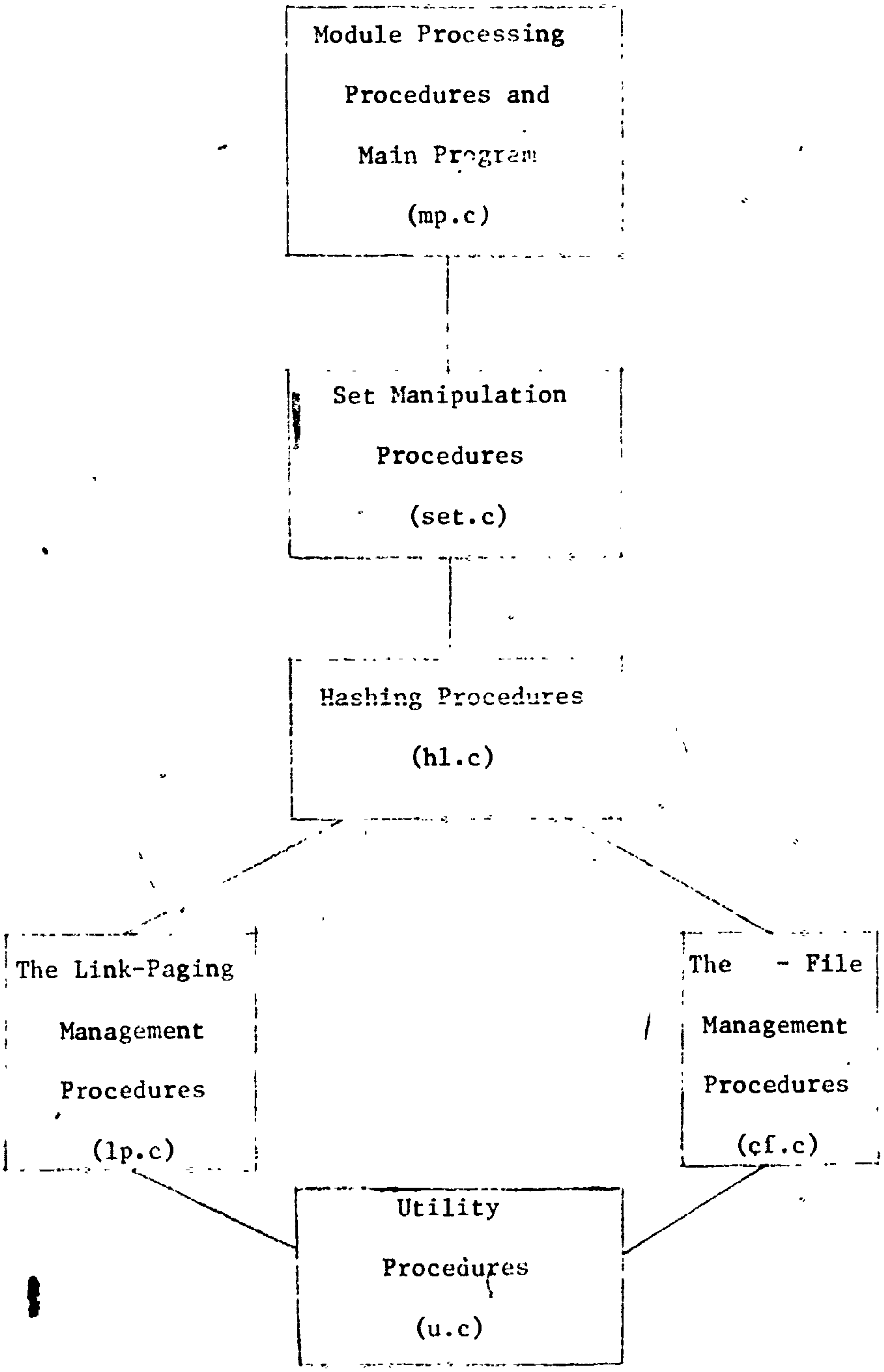


Fig.11

LATTICE DIAGRAM OF PROCEDURES OF LINK-PAGING SYSTEM

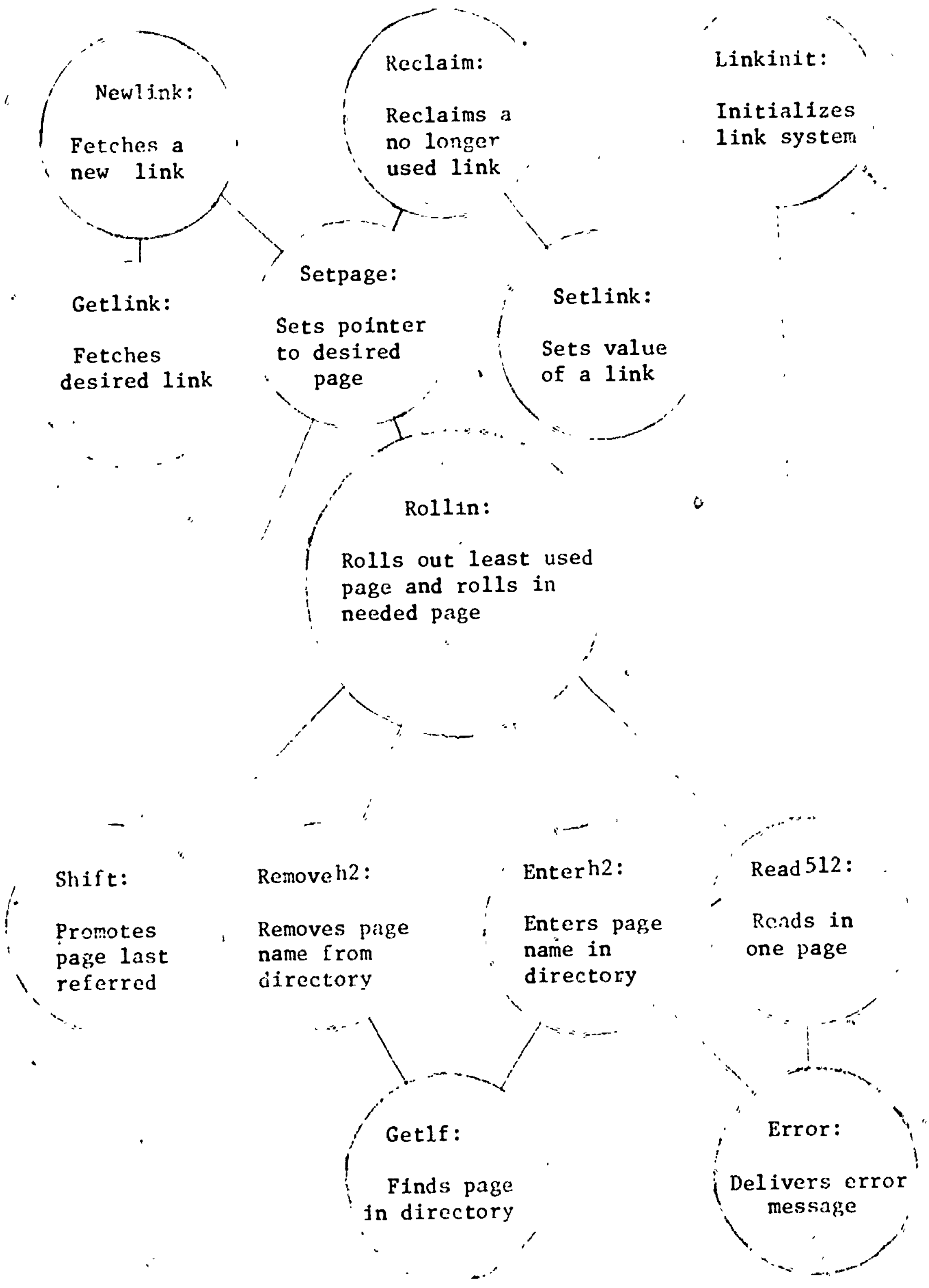


Fig. 12

NEWLINK:
FETCHES A NEW LINK

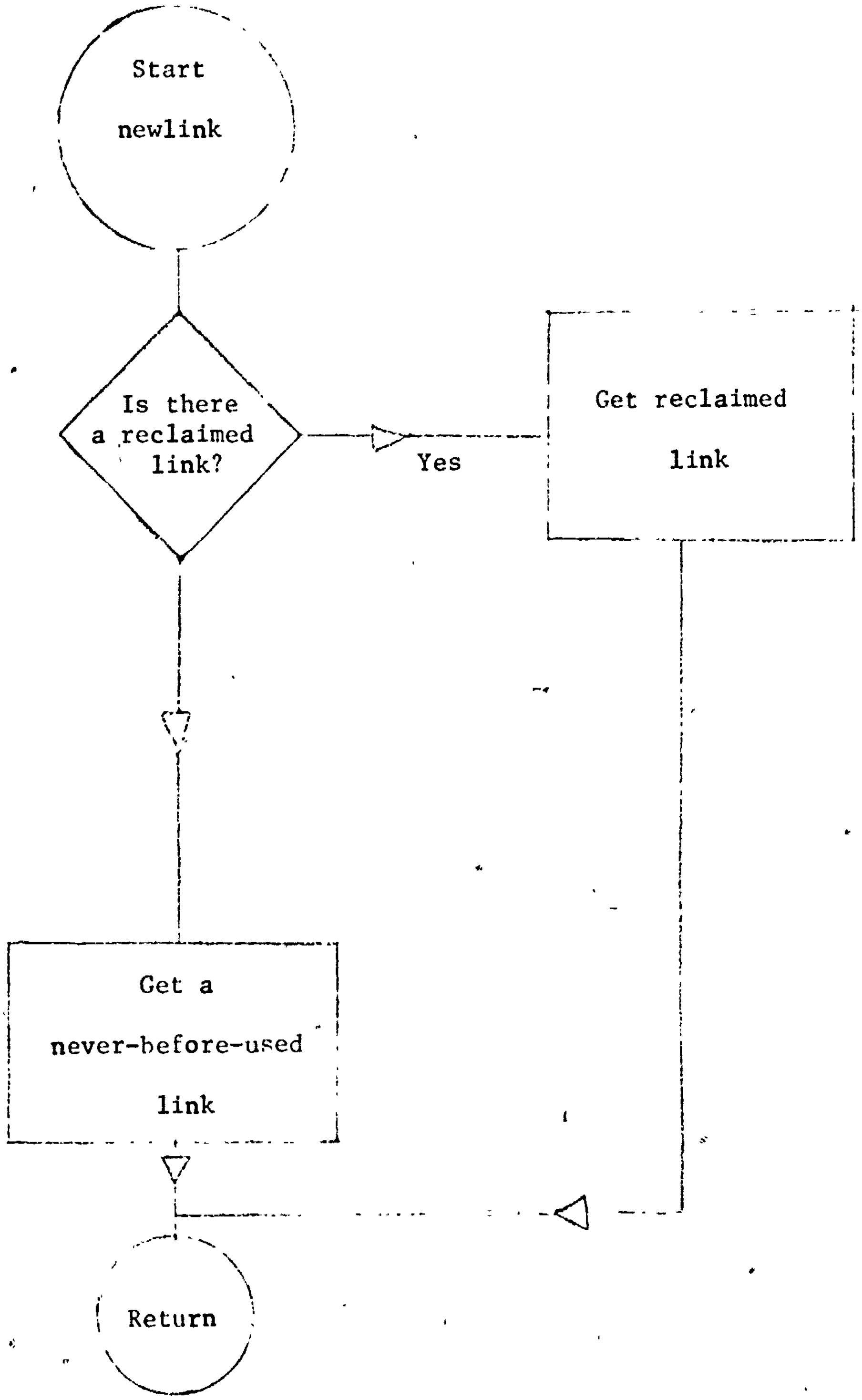


Fig. 13

RECLAIM:
RECLAIMS NO LONGER NEEDED LINK

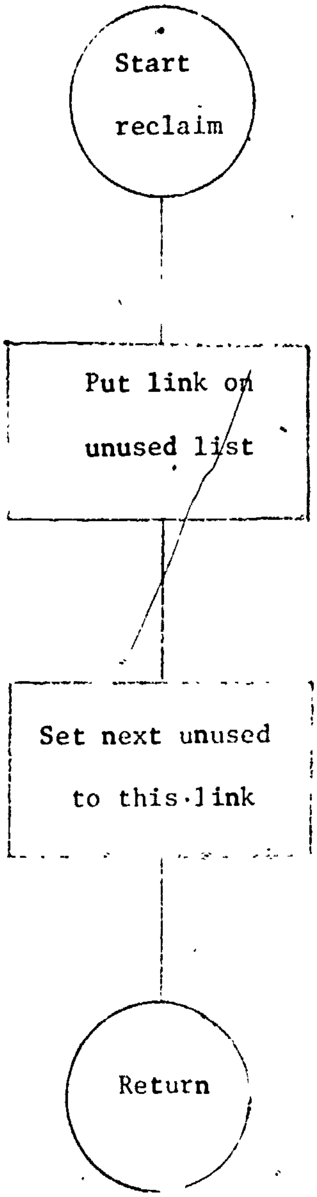


Fig. 14

GETLINK:
FETCHES DESIRED LINK

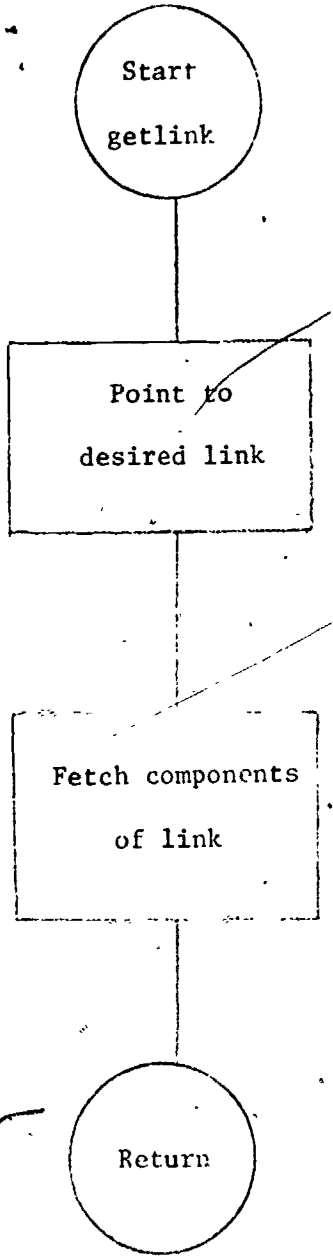


Fig. 15

SETLINK:
SETS NEW VALUES OF A LINK

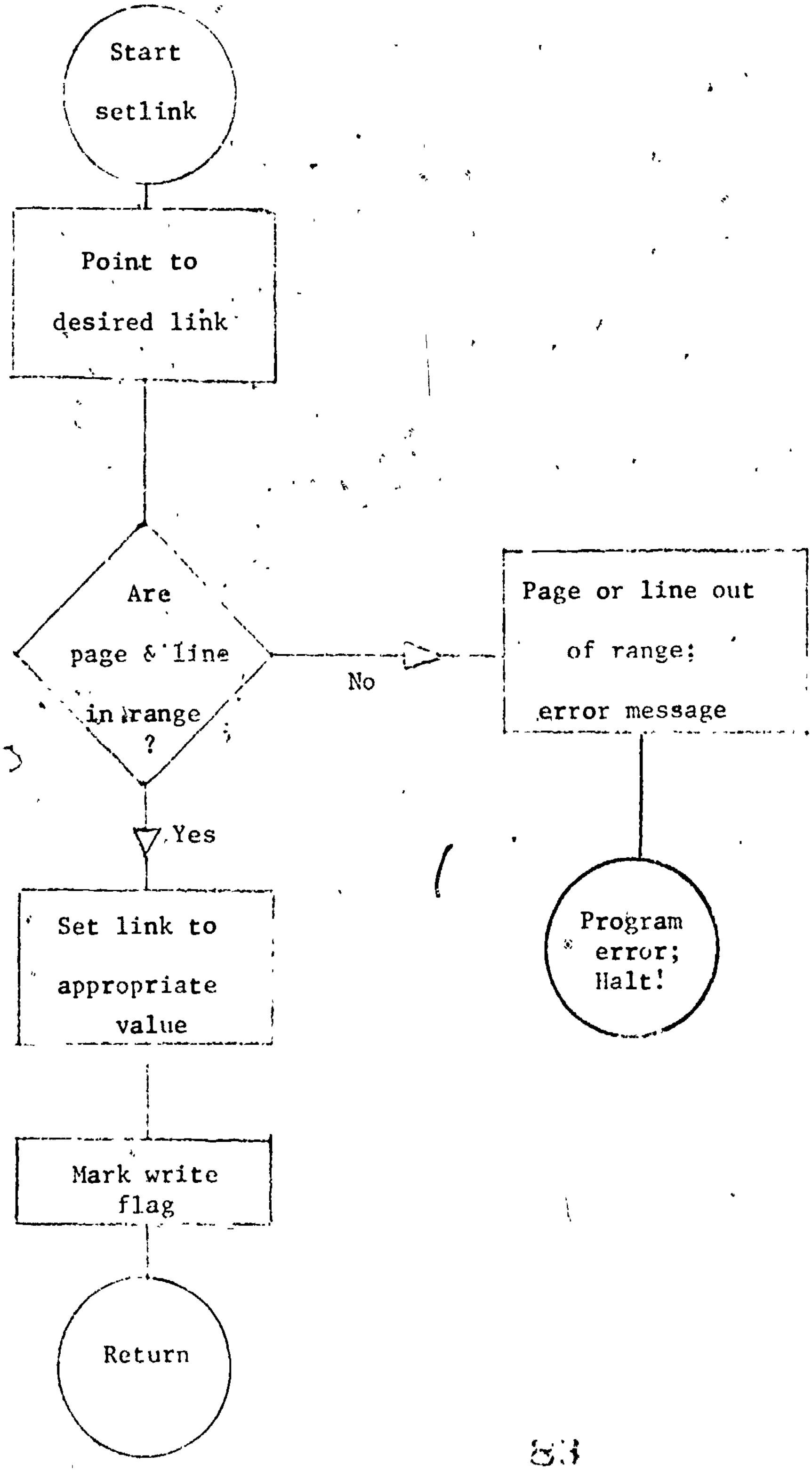


Fig. 16

SETPAGE:
SETS PAGE POINTER TO DESIRED PAGE

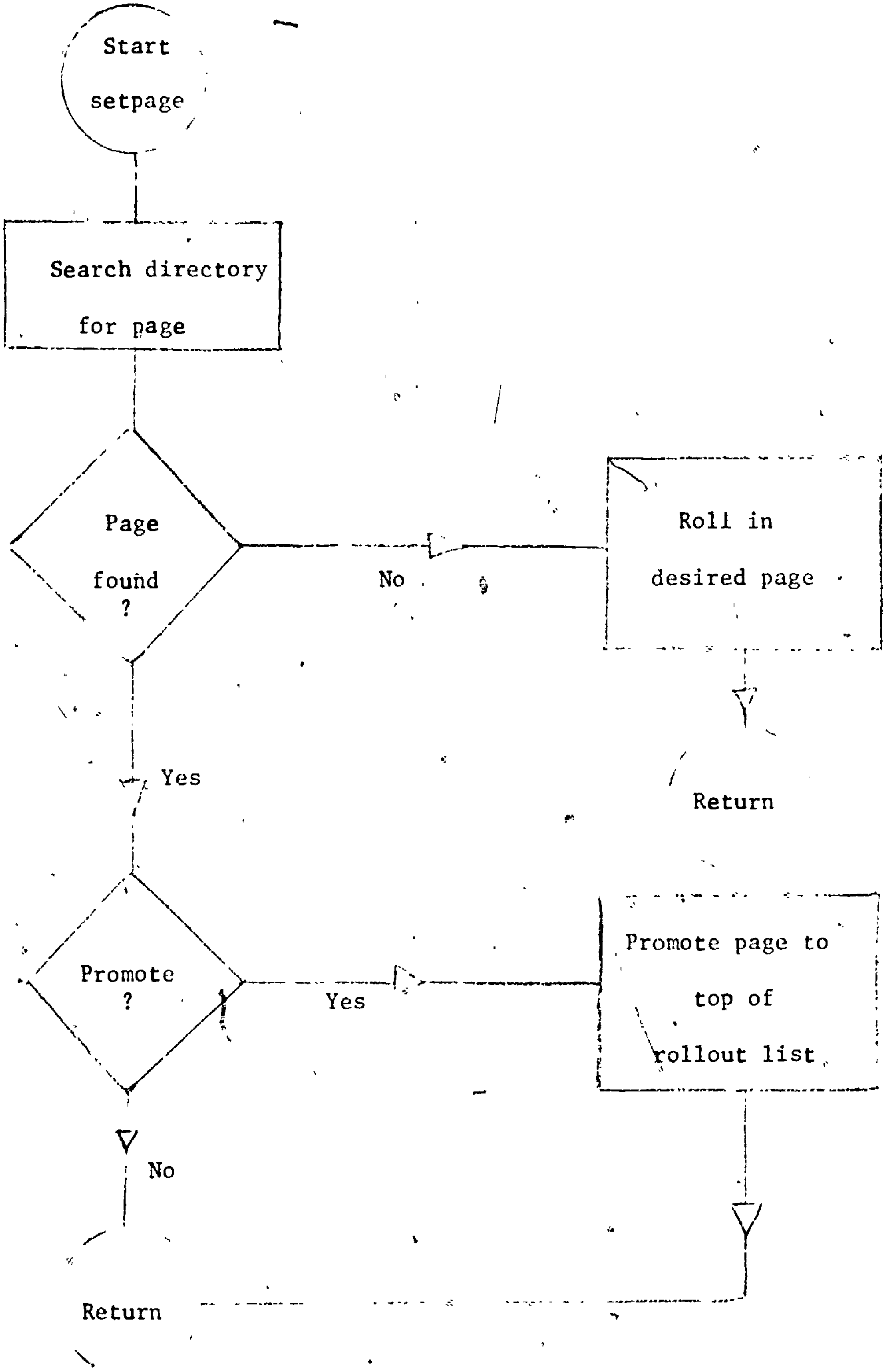


Fig. 17

ROLLIN:
ROLLS OUT OLD PAGE AND BRINGS IN NEWS

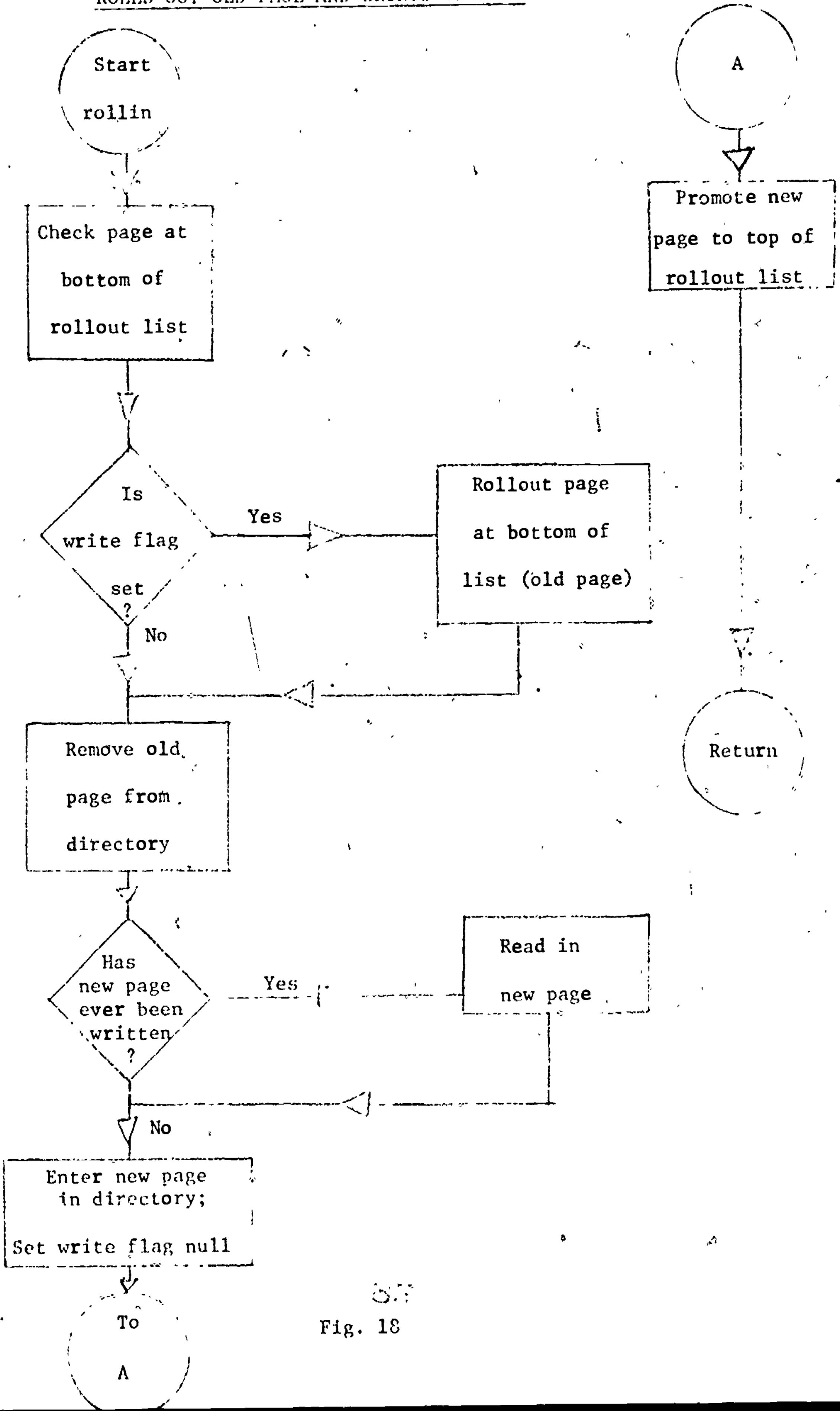


Fig. 18

ENTERH2:
ENTER NEW PAGE NAME IN DIRECTORY

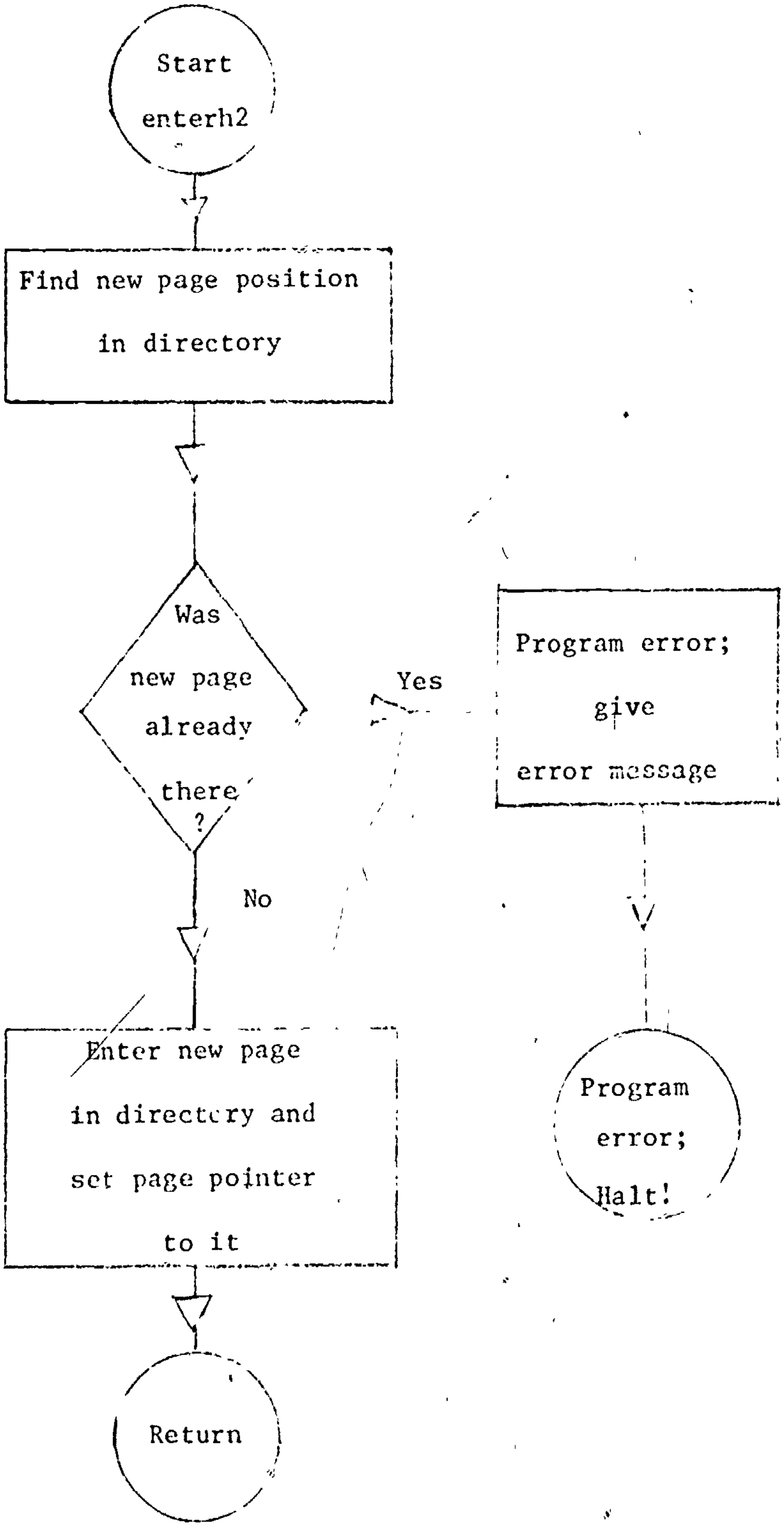


Fig. 19

REMOVEH2:
REMOVES OLD PAGE NAME FROM DIRECTORY

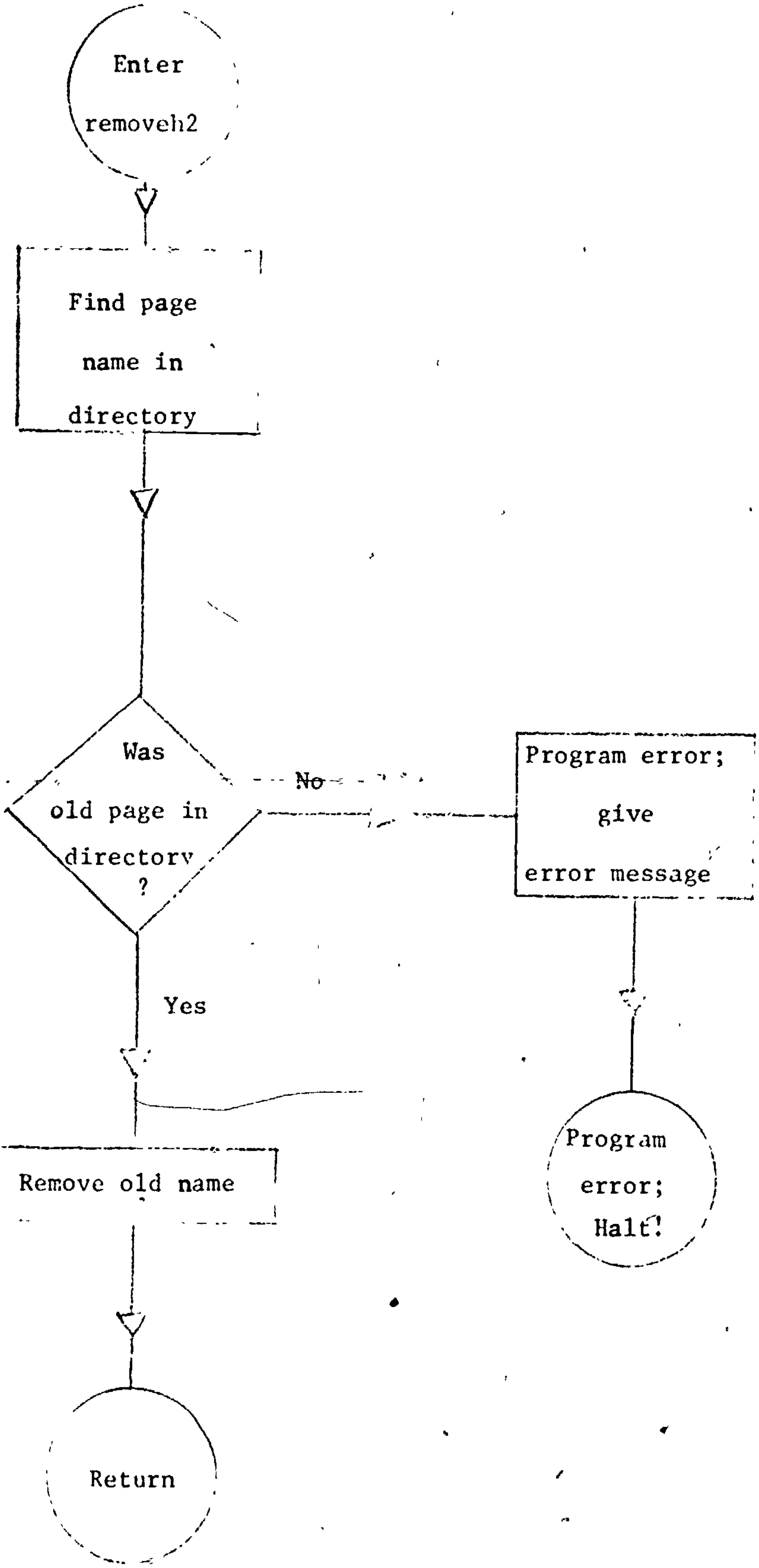


Fig. 20

GETLF:
FIND PAGE NAME OR POSITION IN DIRECTORY

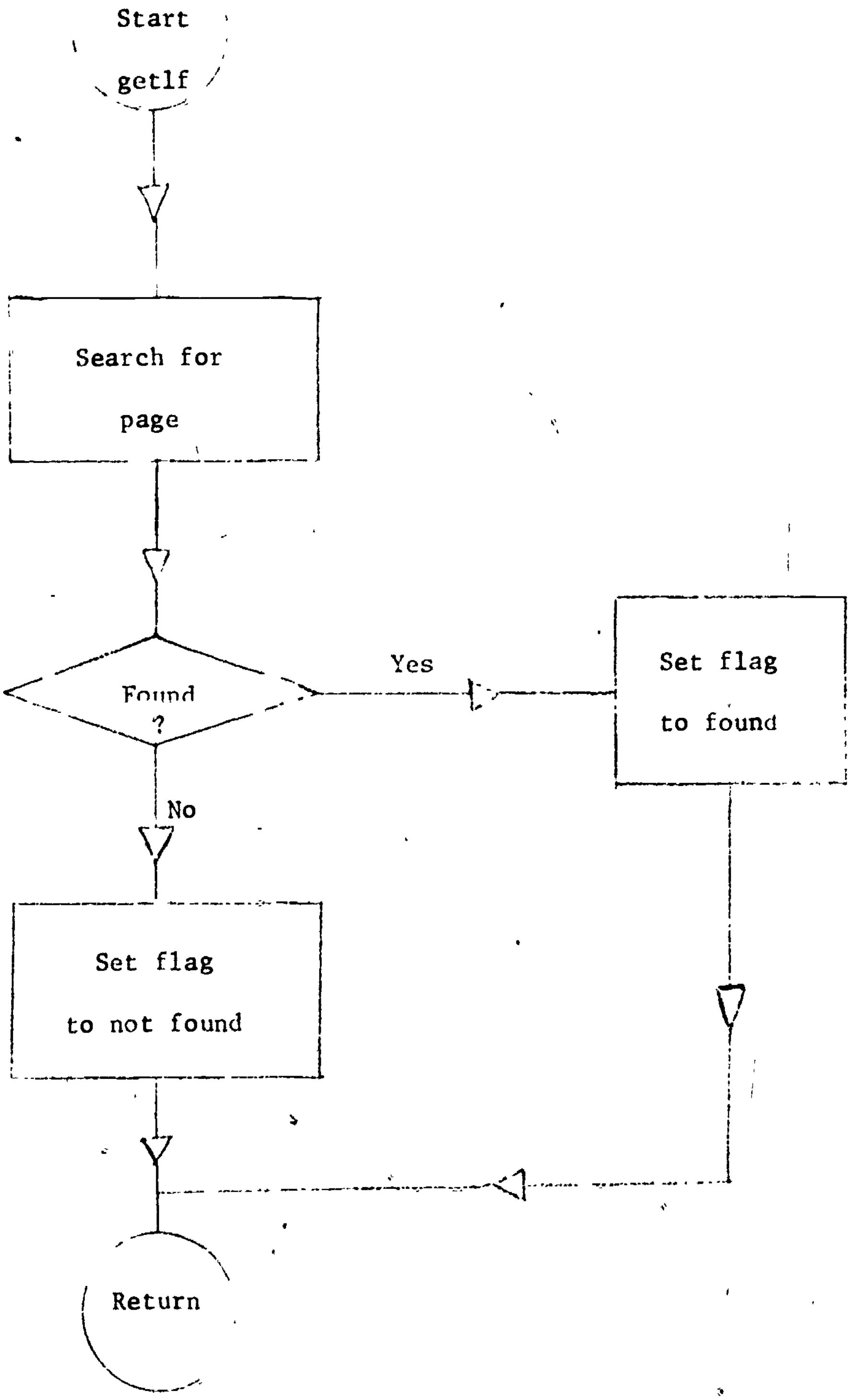


Fig. 21

SHIFT:
PROMOTE PAGE LAST REFERENCED TO TOP OF ROLL OUT LIST

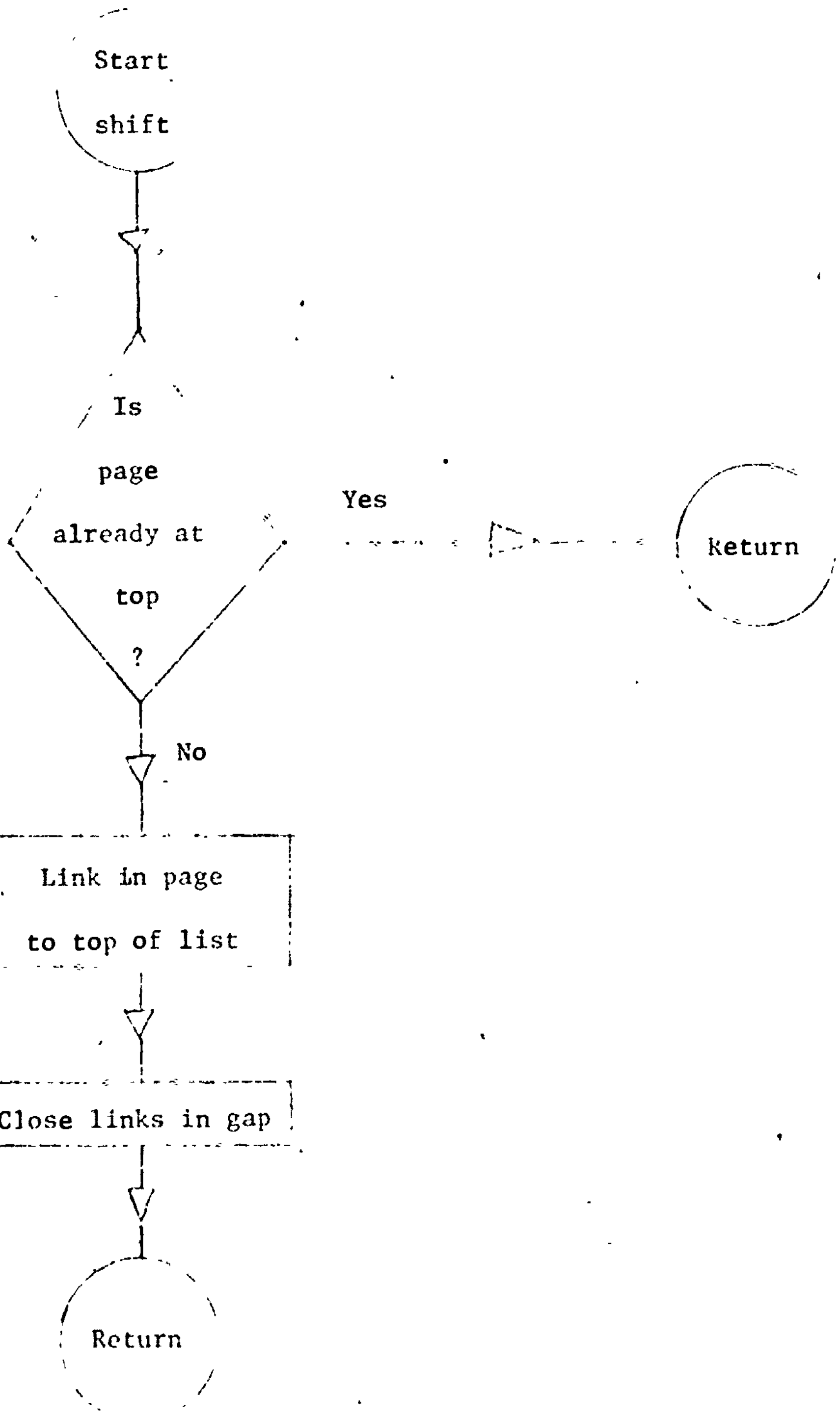


Fig. 22

LINKINIT: INITIALIZES LINKAGE SYSTEM

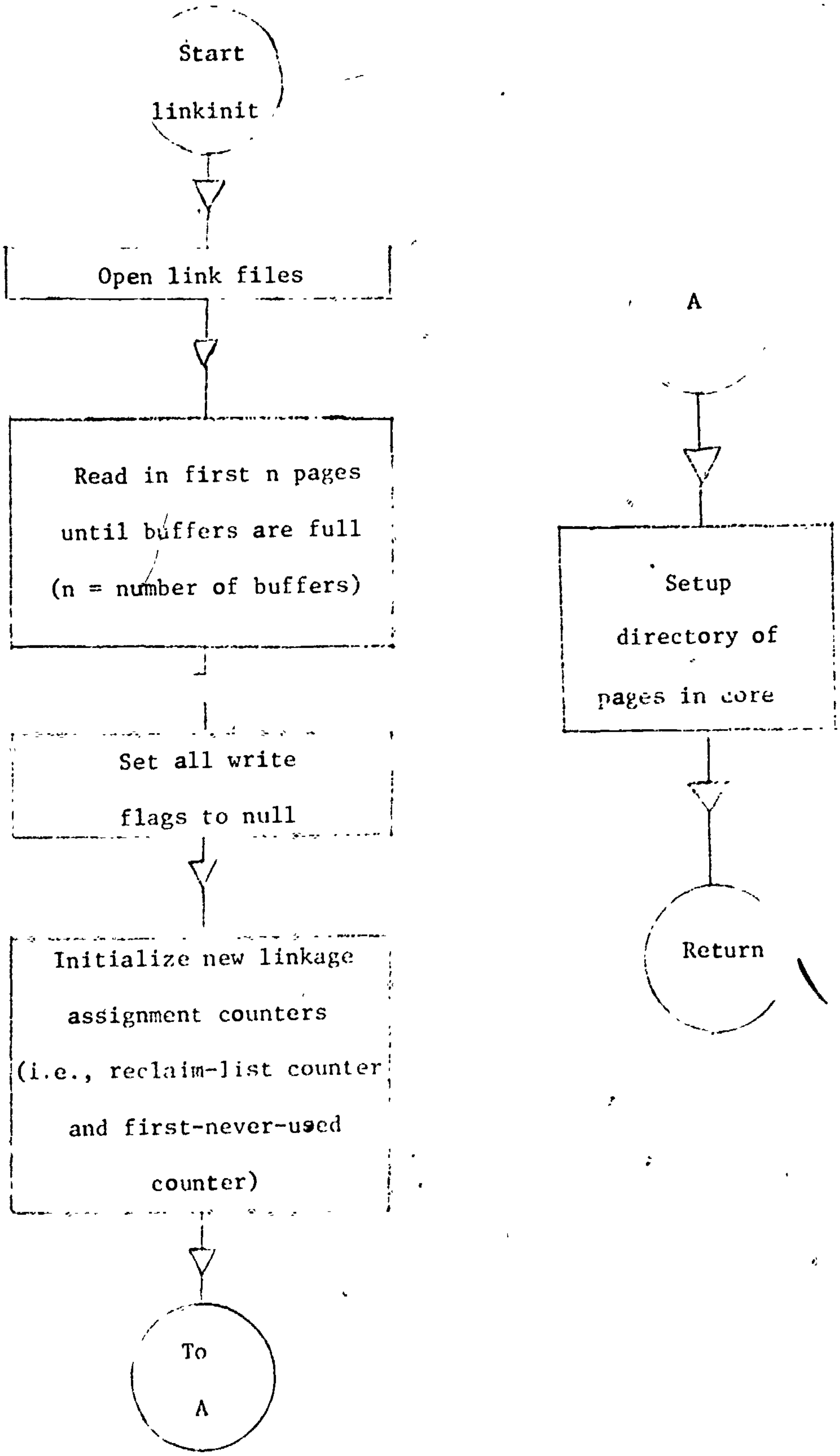


Fig. 23

READ512:
READS IN ONE PAGE (512 CHARACTERS) OF LINKS

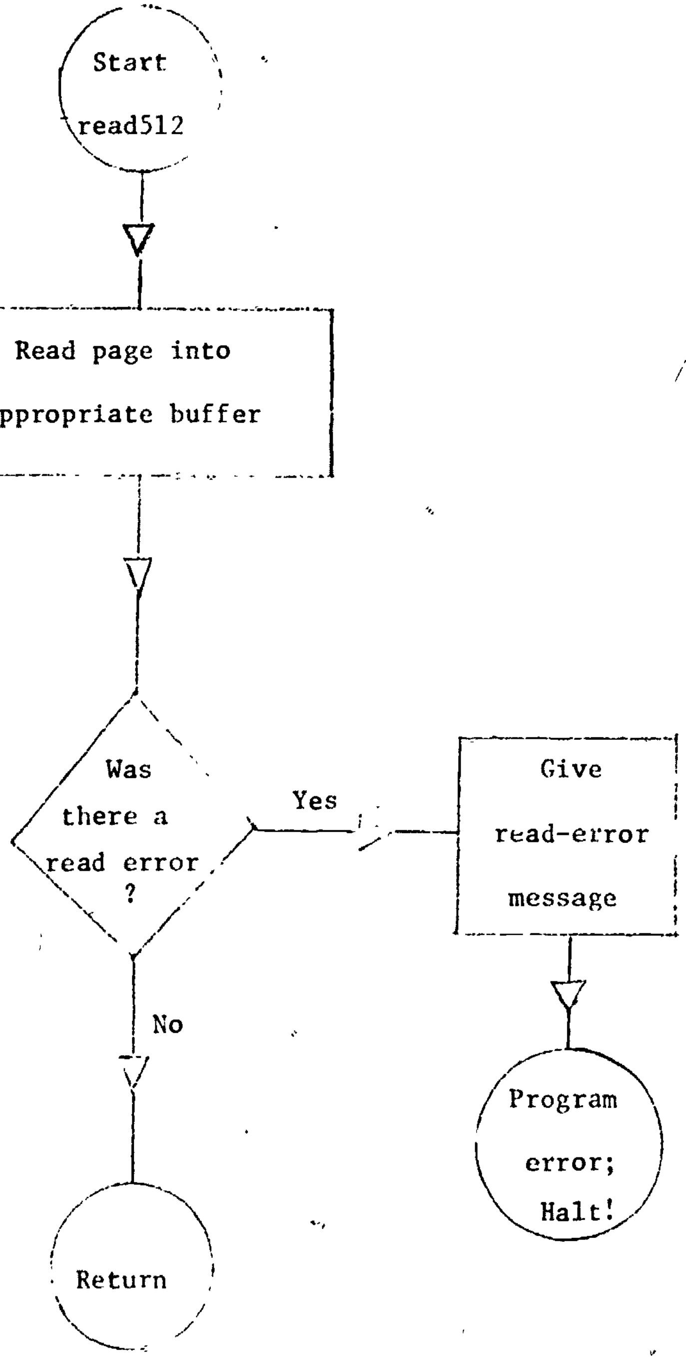


Fig. 24

LATTICE DIAGRAM OF α -FILE MANAGEMENT BLOCK (cf.c)

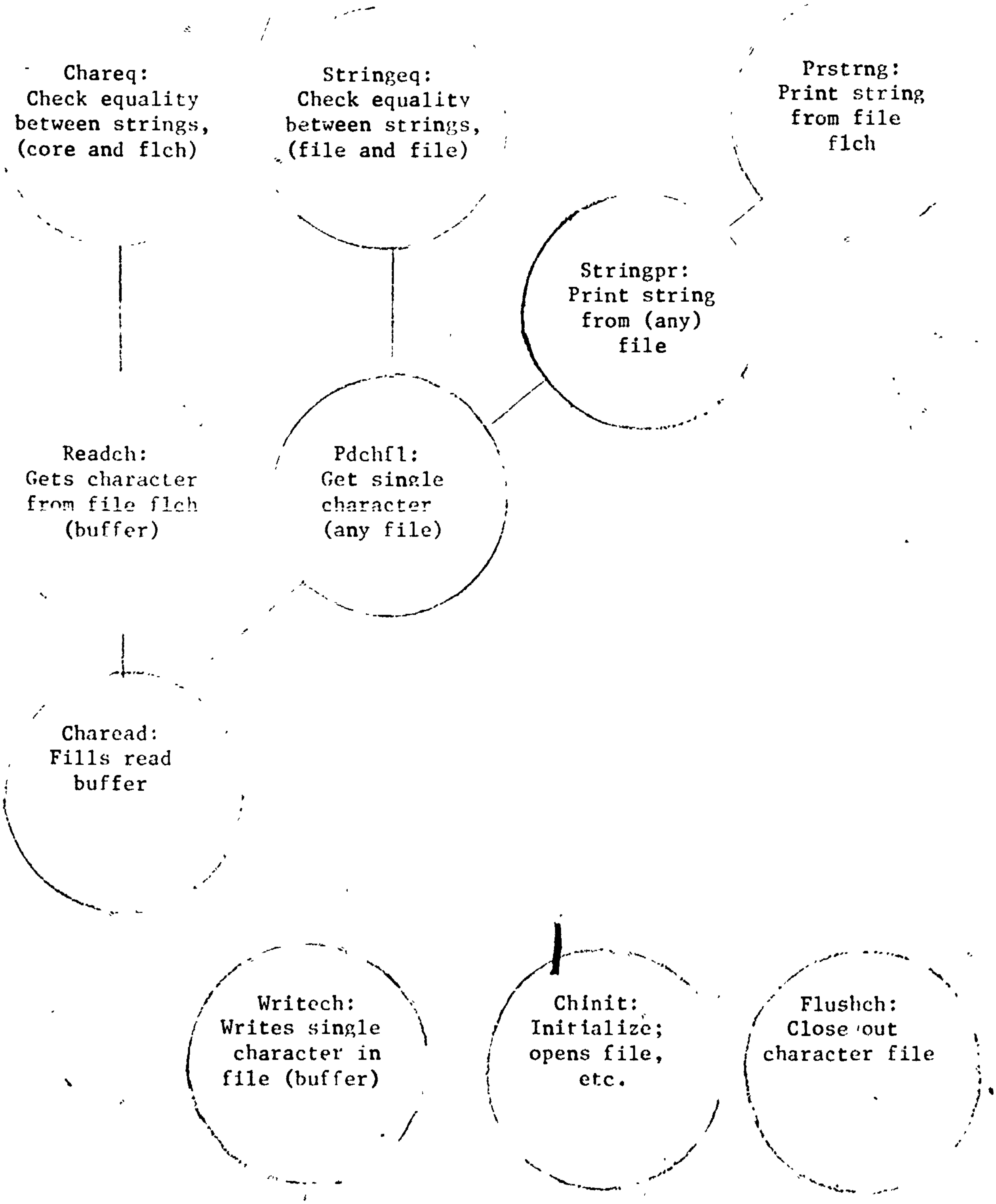


Fig. 25

WRITECH:
WRITES A CHARACTER ON FILE (BUFFER)

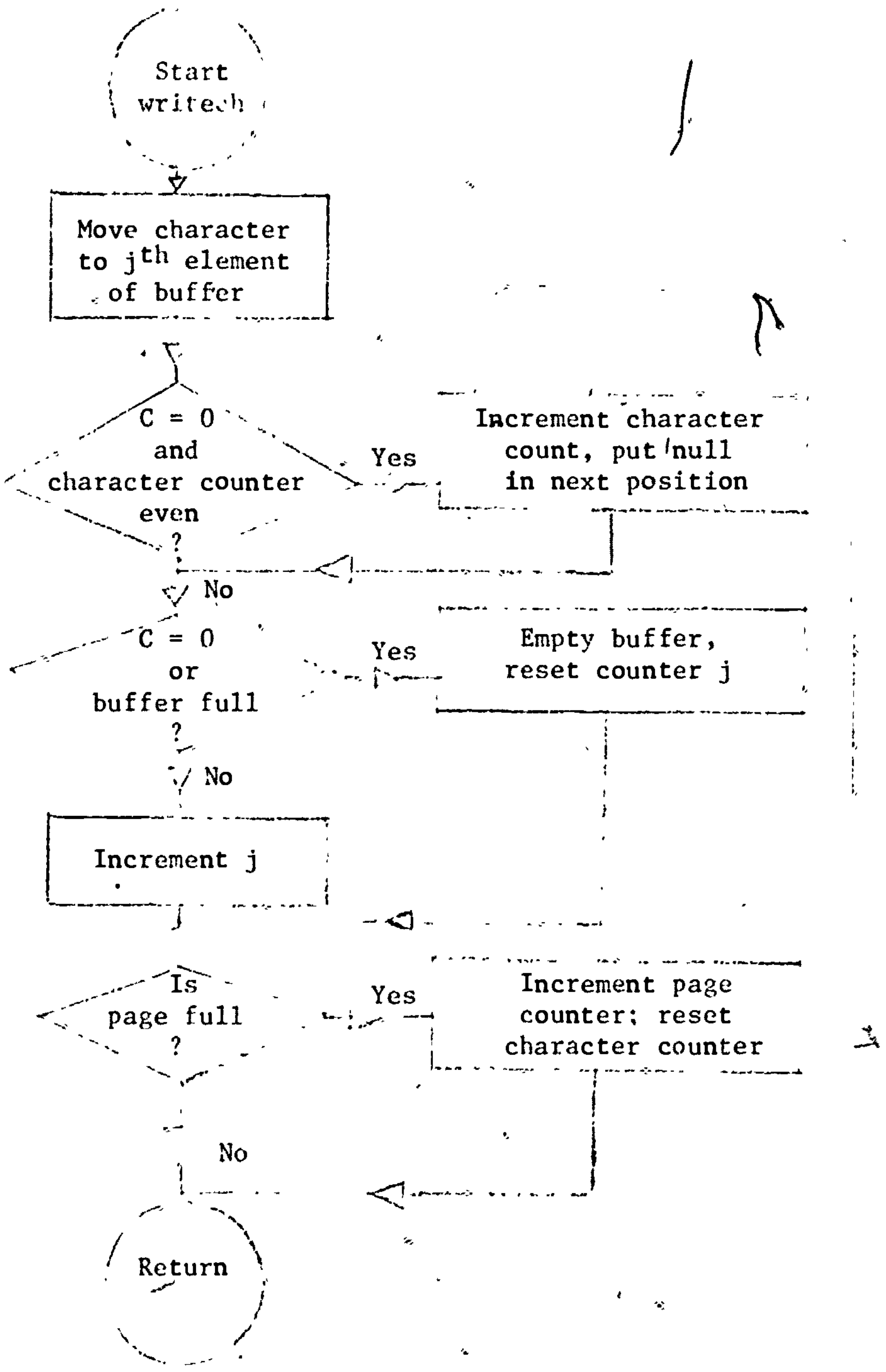


Fig. 26

CHAREAD:
FILLS READ BUFFER FROM FILE

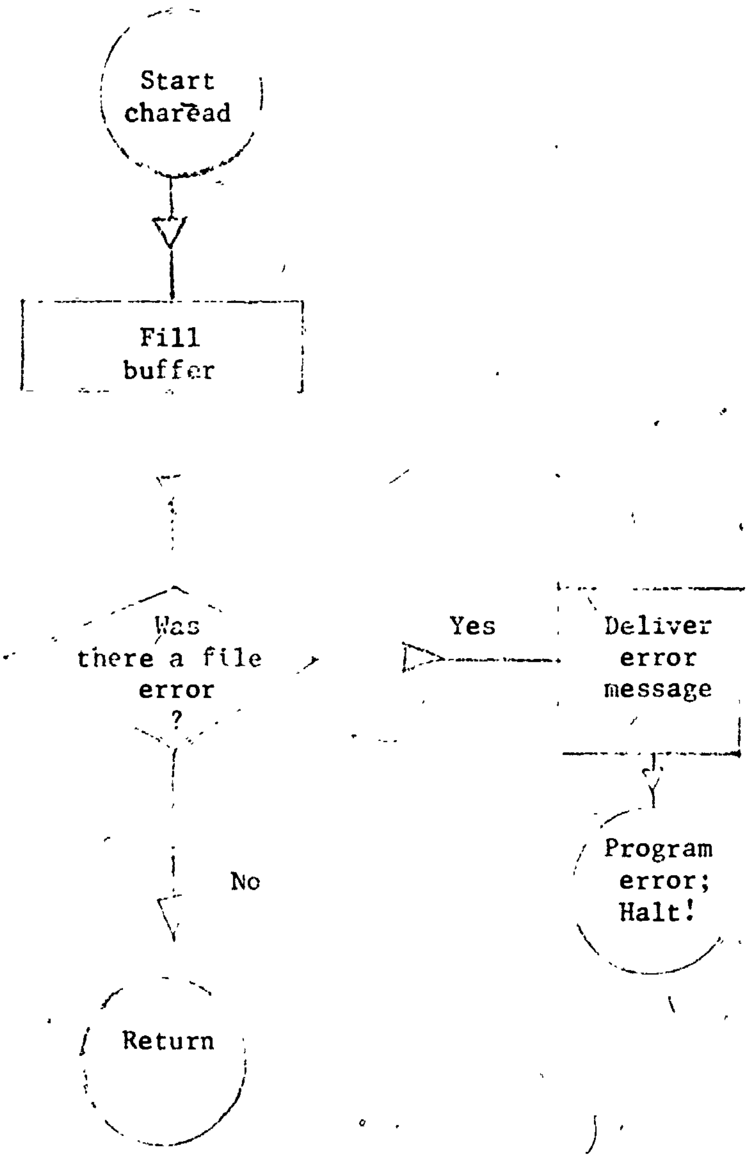


Fig. 27

RDCHFC:
READS A SINGLE CHARACTER FROM A FILE (BUFFER)

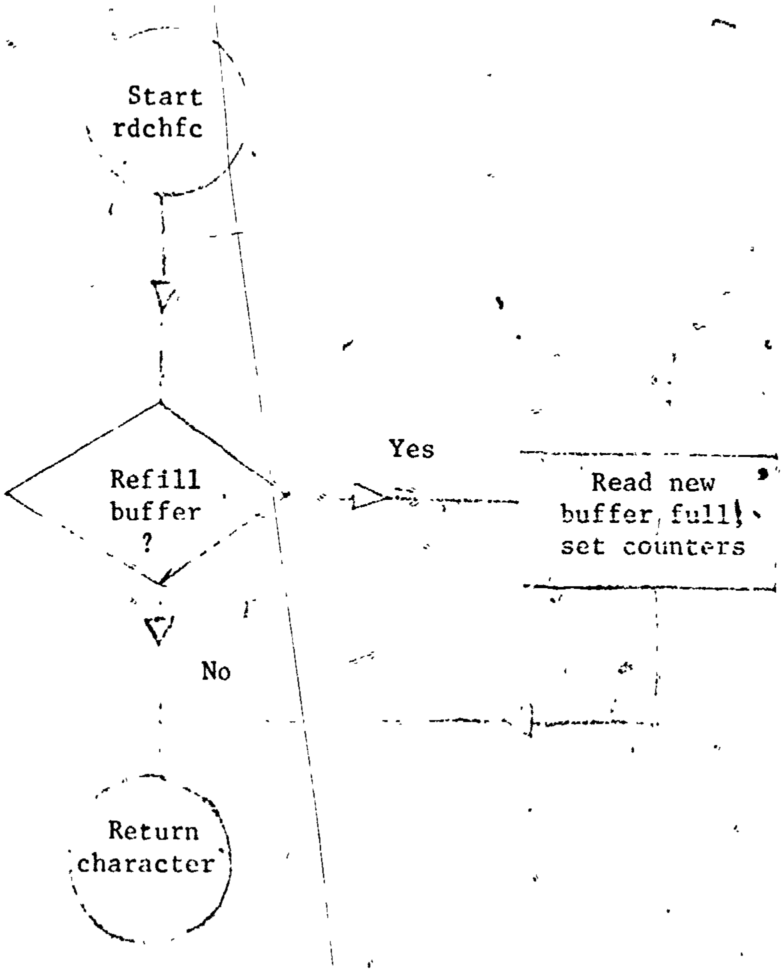


Fig. 28

STRINGPR:
PRINTS STRING FROM FILE

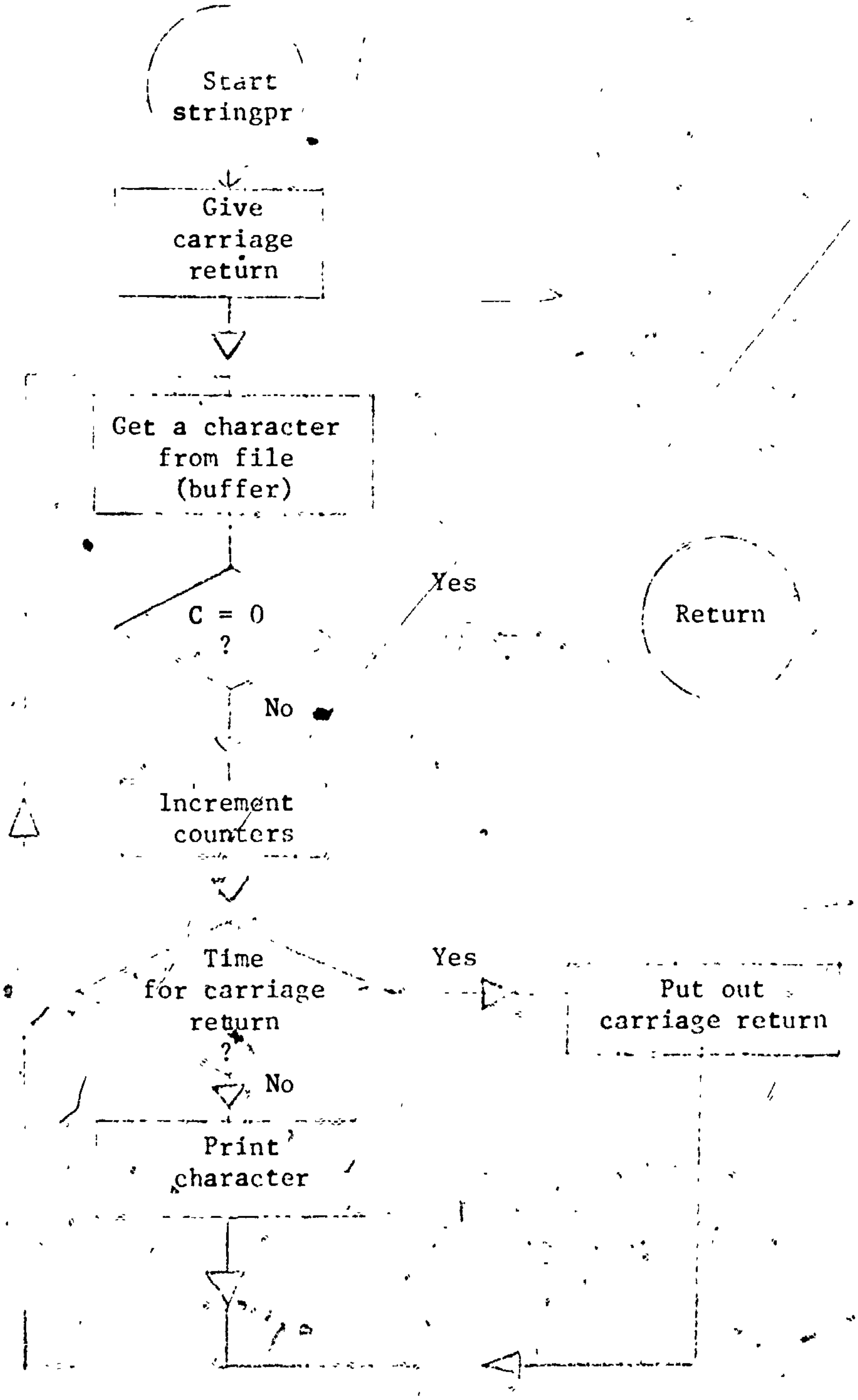


Fig. 29

PRSTRING:
PRINT STRING FROM FILE FLCH

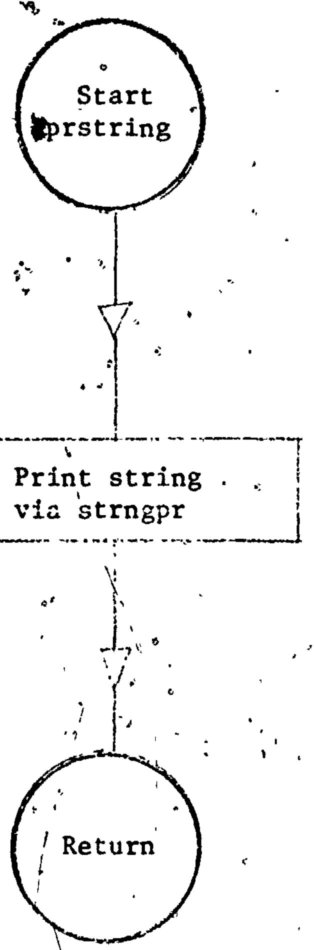


Fig. 30

STRNGEQ:

TESTS FOR EQUALITY BETWEEN TWO STRINGS IN (POSSIBLY DIFFERENT) FILES

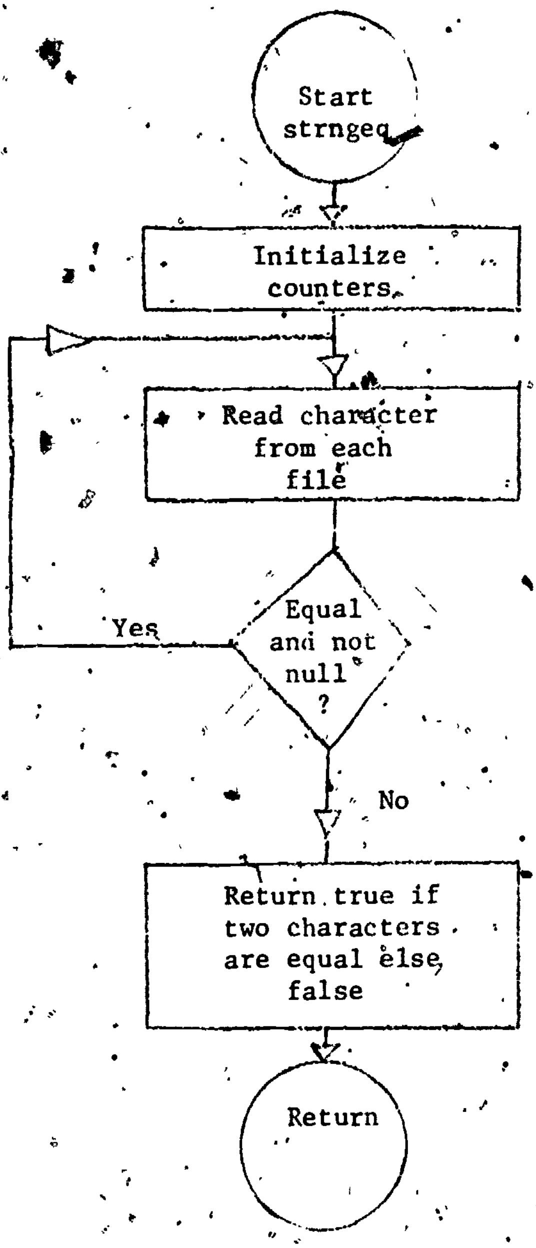


Fig. 31

READCH:
READS A SINGLE CHARACTER FROM FILE FLCH

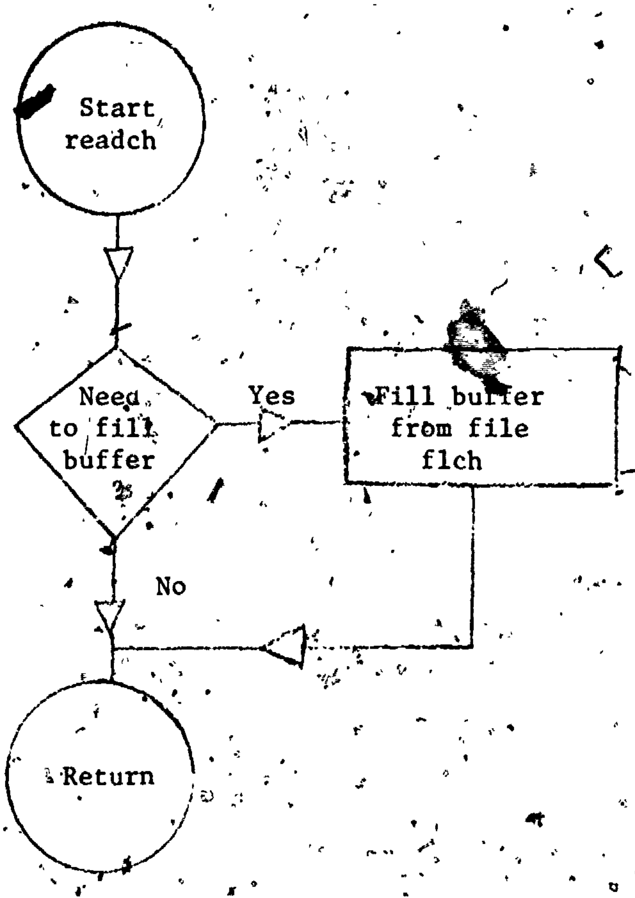


Fig. 32

CHAREQ:
EQUALITY TEST BETWEEN A STRING IN CORE AND A STRING ON FILE

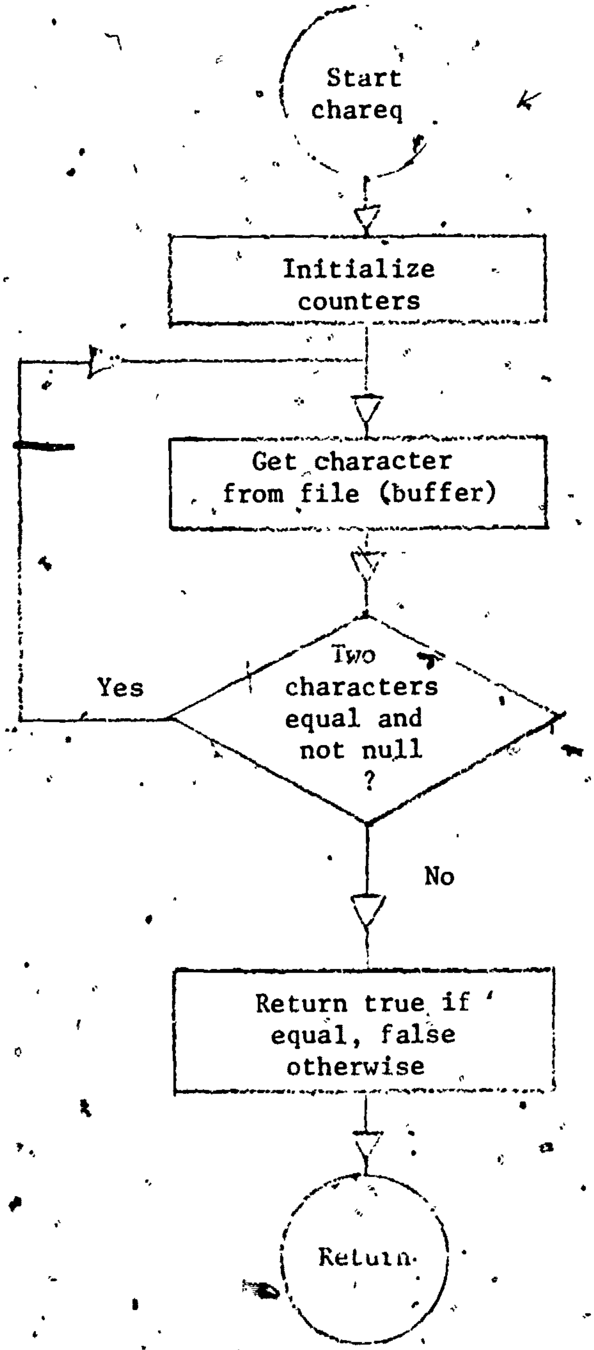


Fig. 33

CHINIT:
INITIALIZE, OPEN FILES, ETC.

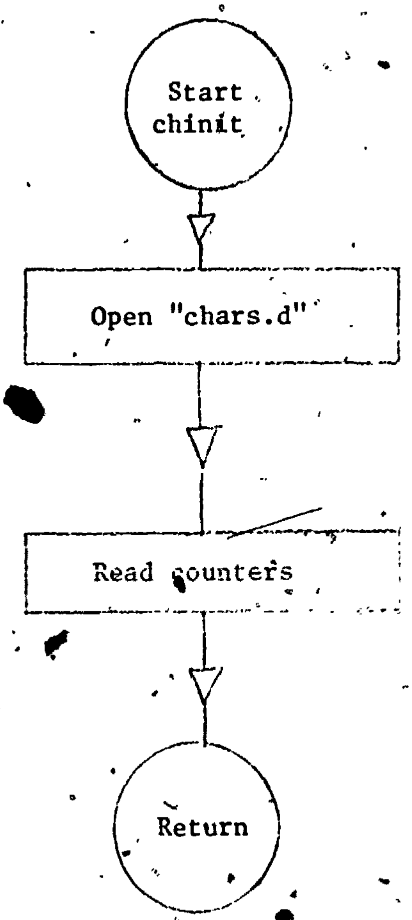


Fig. 34

FLUSHCH:
CLOSES OUT THE CHARACTER FILE

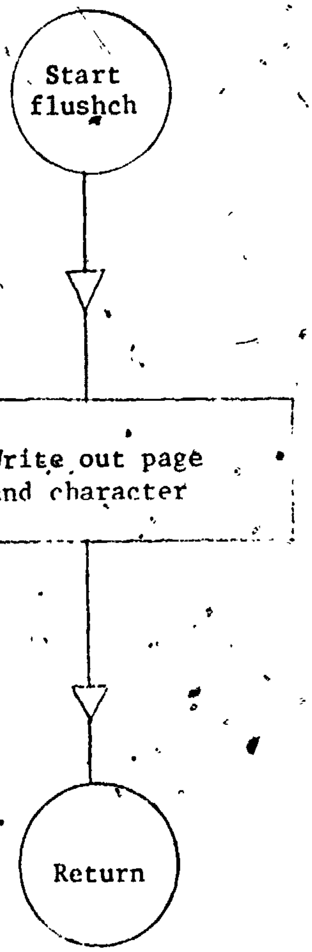


Fig. 35

BALLOON DIAGRAM:
HASH (01) ALGARITHM
BLOCK* (h1.c)

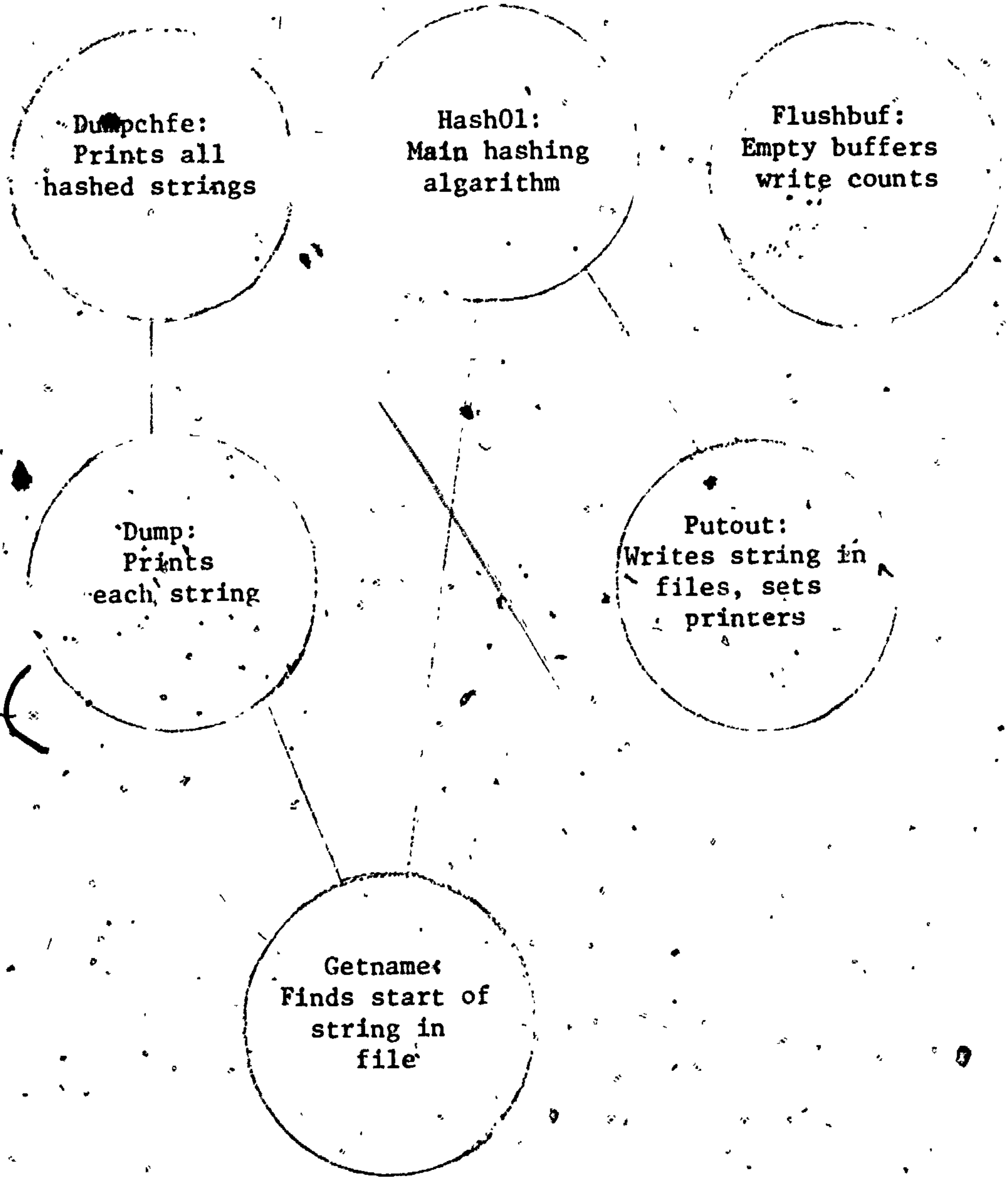


Fig. 36

FLUSHBUF:
EMPTY BUFFERS, WRITE COUNTERS

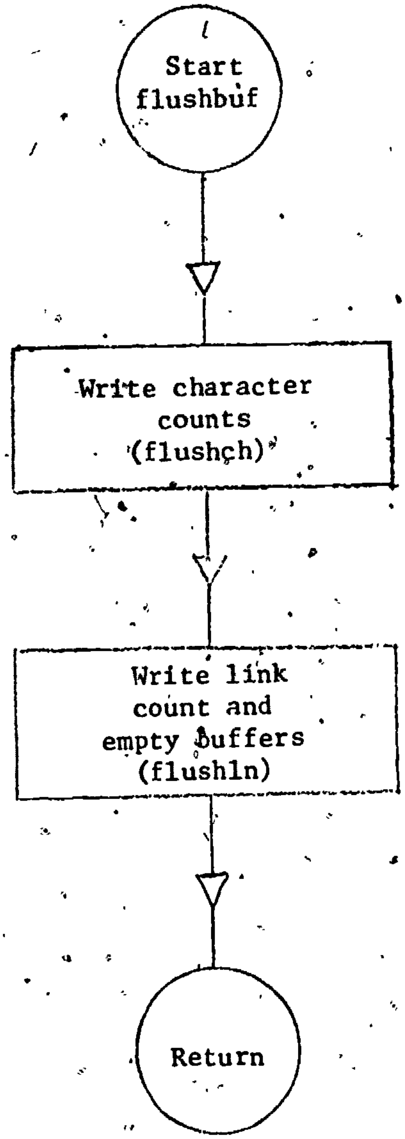


Fig. 37

DMPCHFL:

WRITES OUT ALL THE STRINGS HASHED INTO THE CHARACTER FILE

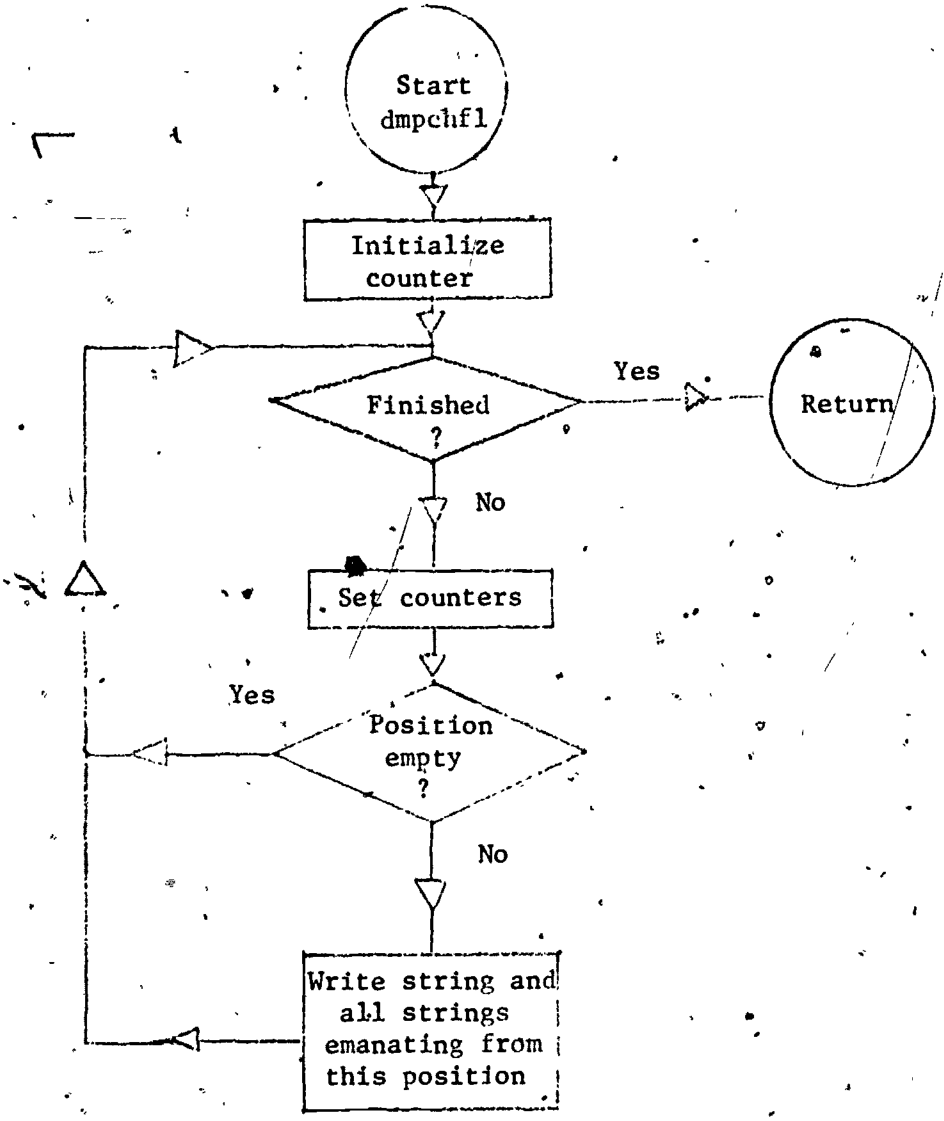


Fig. 38

DUMP:
WRITE OUT A PARTICULAR STRING AND ALL STRINGS EMANATING FROM THAT ONE

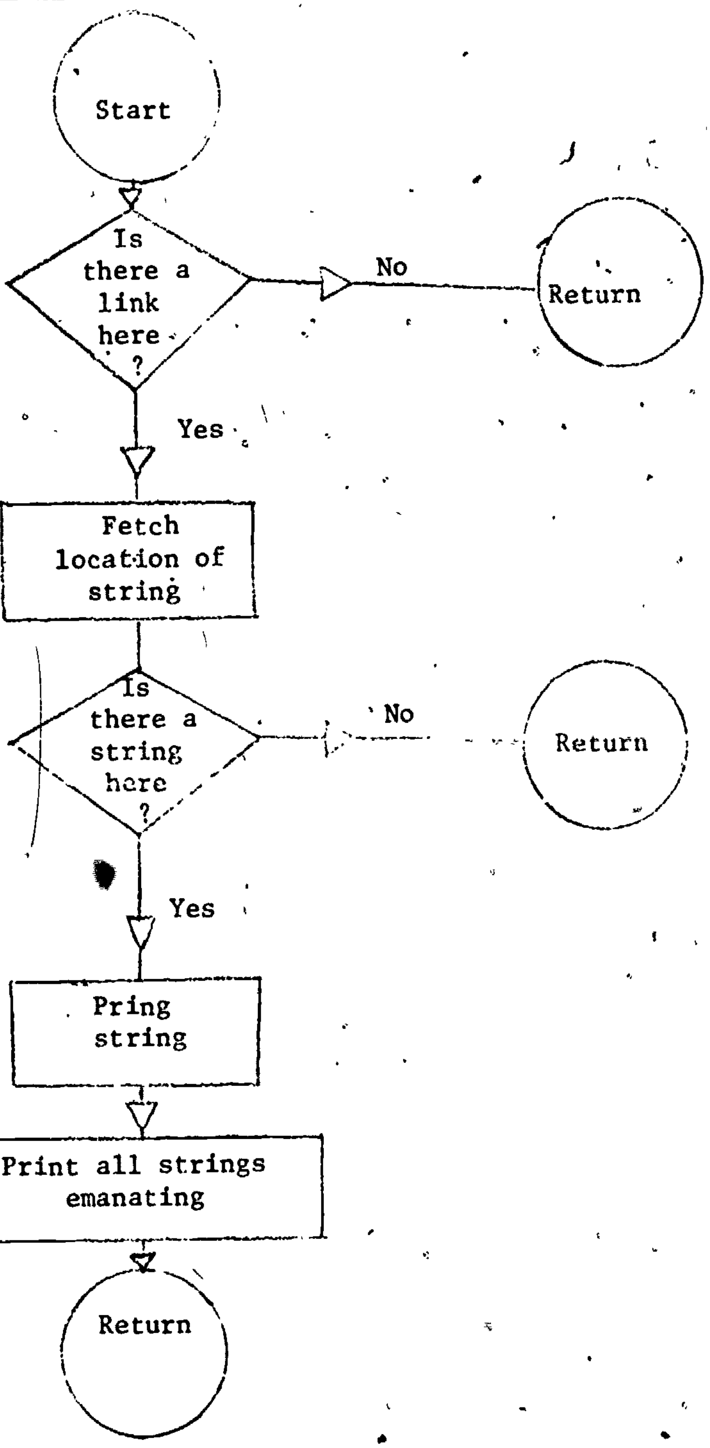


Fig. 39

HASH01: MAIN HASHING ALGORITHM

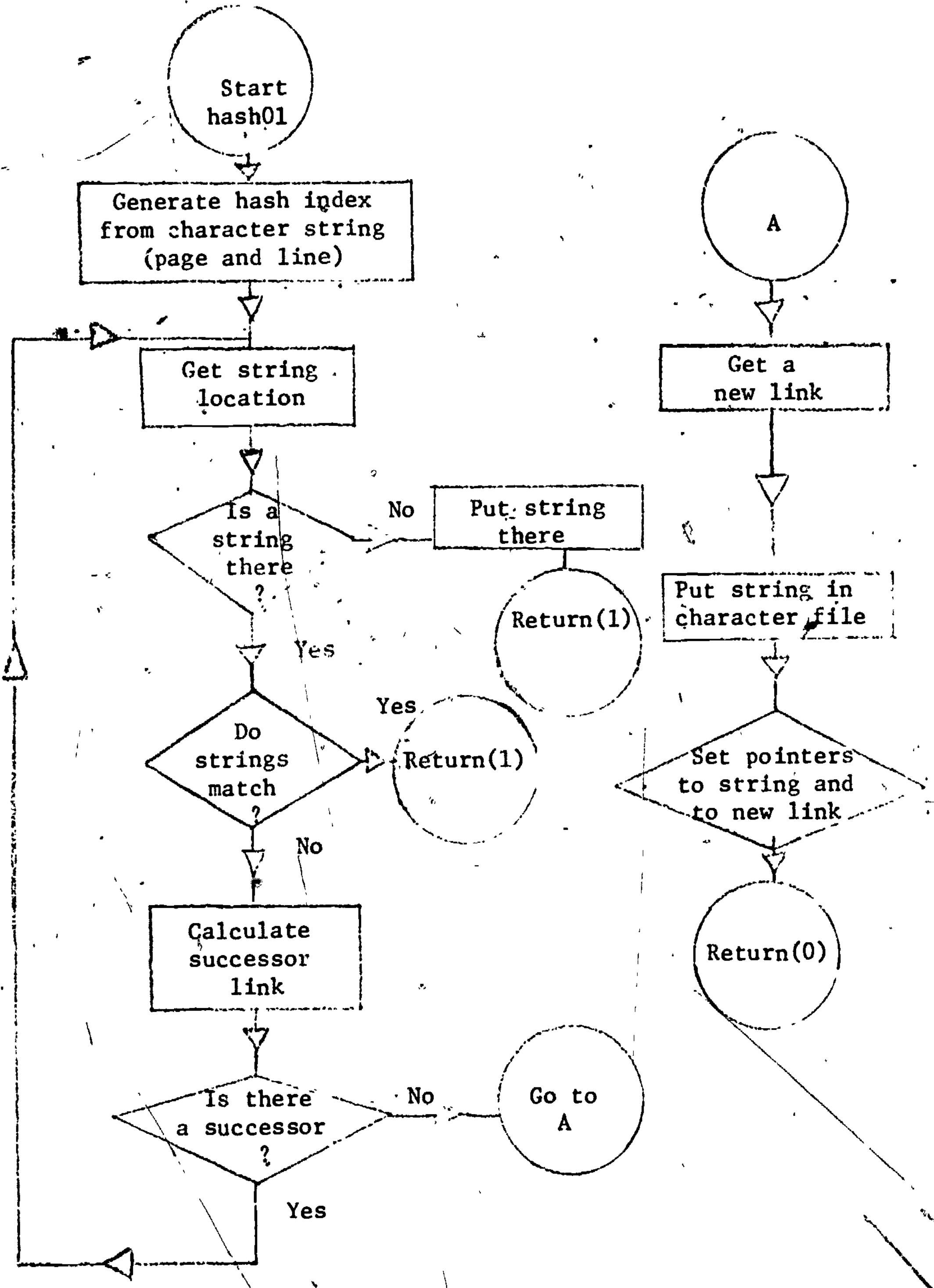


Fig. 40

PUTAT:
WRITES STRING, SETS POINTERS TO IT

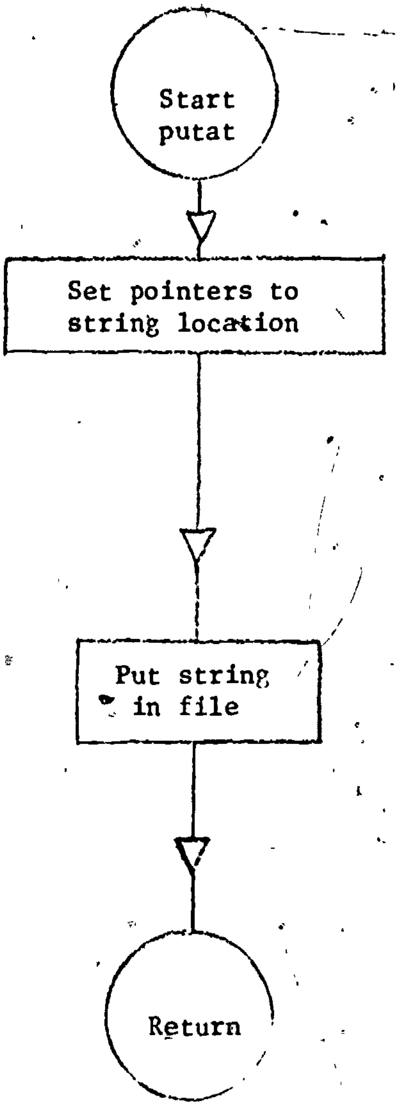


Fig. 41

GETNAME:
FINDS START OF STRING IN CHARACTER FILE

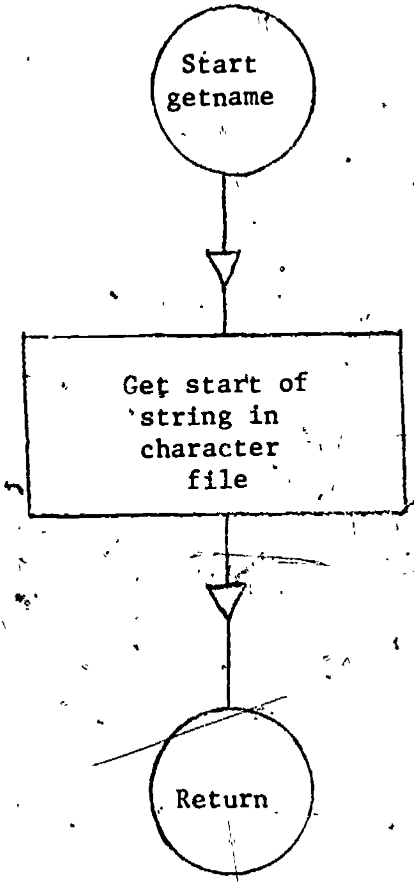
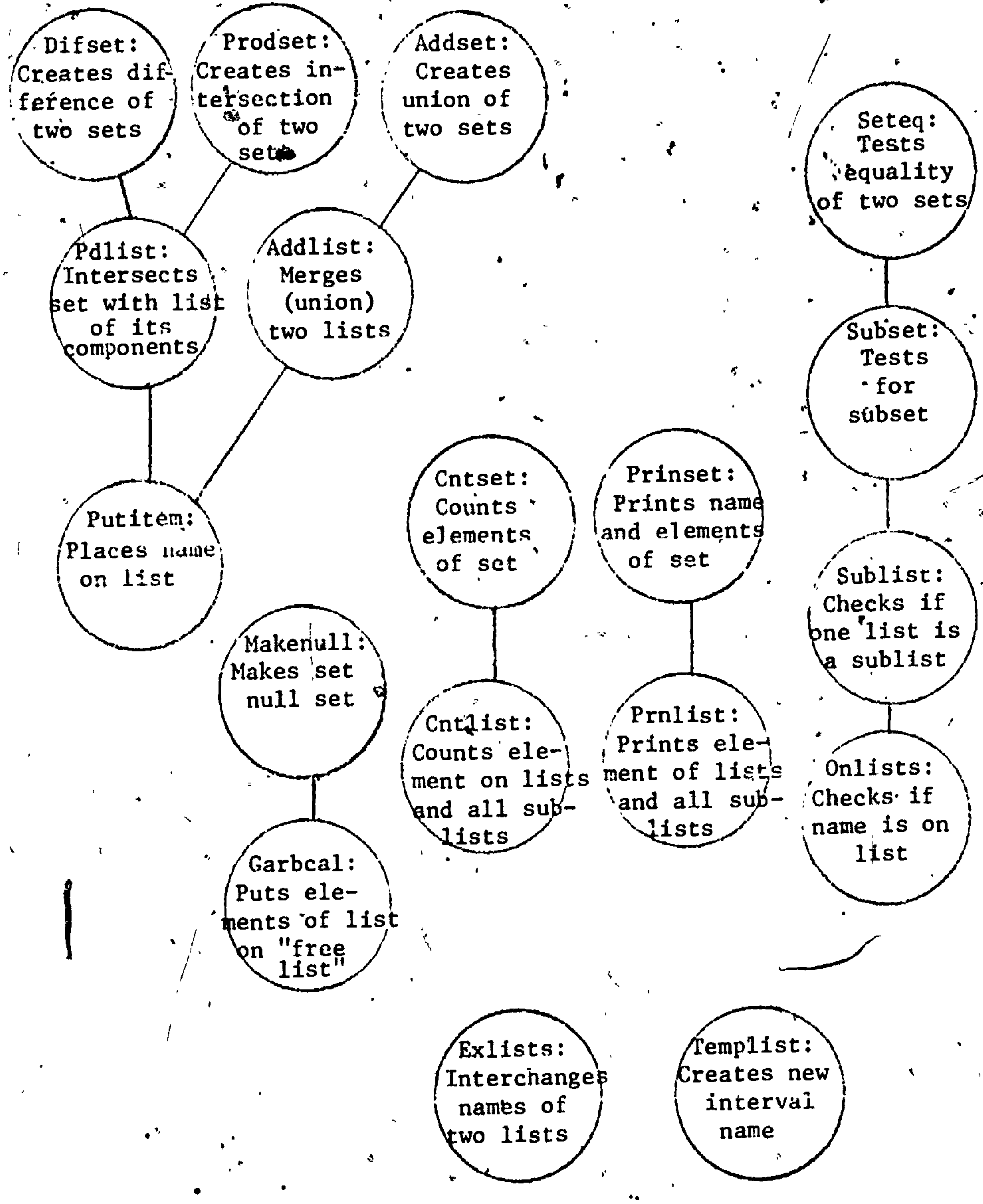


Fig. 42

LATTICE DIAGRAMS FOR PROCEDURES IN SET MANIPULATION (BLOCK* (set.c)



*Procedures called in lower blocks not shown.

Fig. 43

TEMPLIST:
CREATES A LITERAL (EXTERNAL) NAME FOR INTERNAL USE ONLY

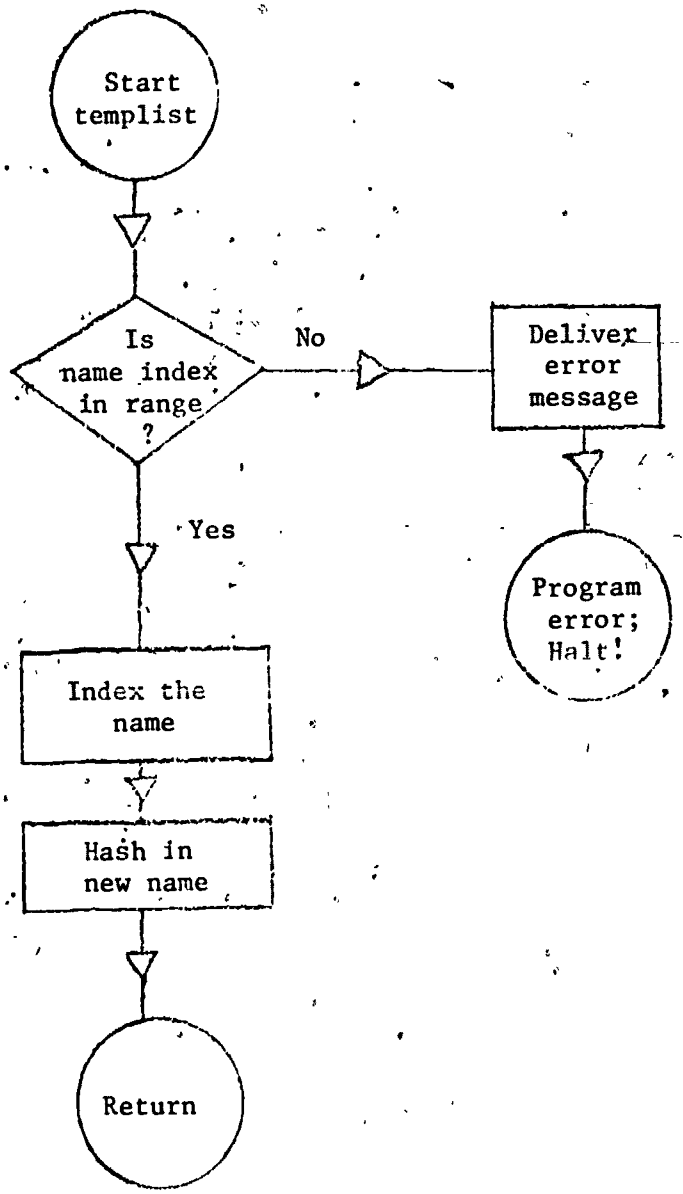


Fig. 44

EXLISTS:
INTERCHANGE NAMES OF TWO LISTS

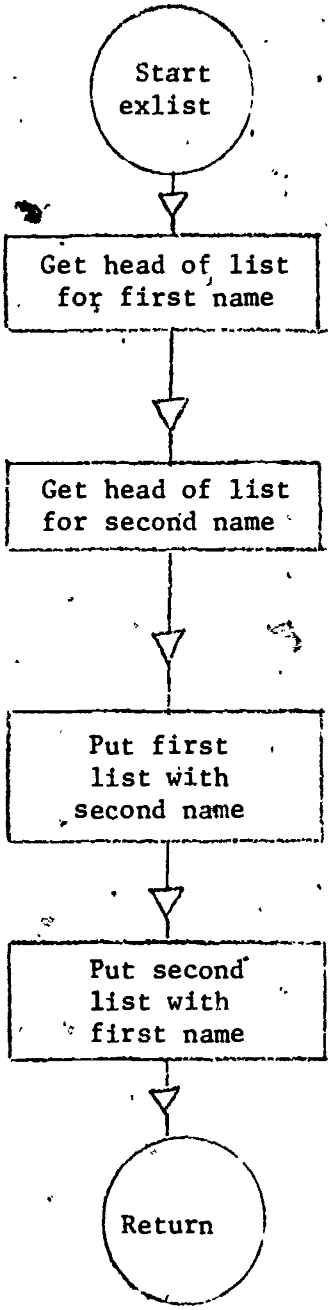


Fig. 45

CNTSET:
COUNTS THE ELEMENTS OF A SET

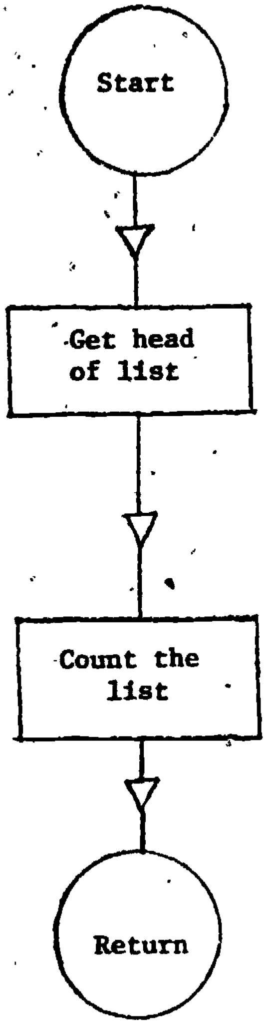


Fig. 46

CNTLIST:
COUNTS THE ITEMS ON A LIST

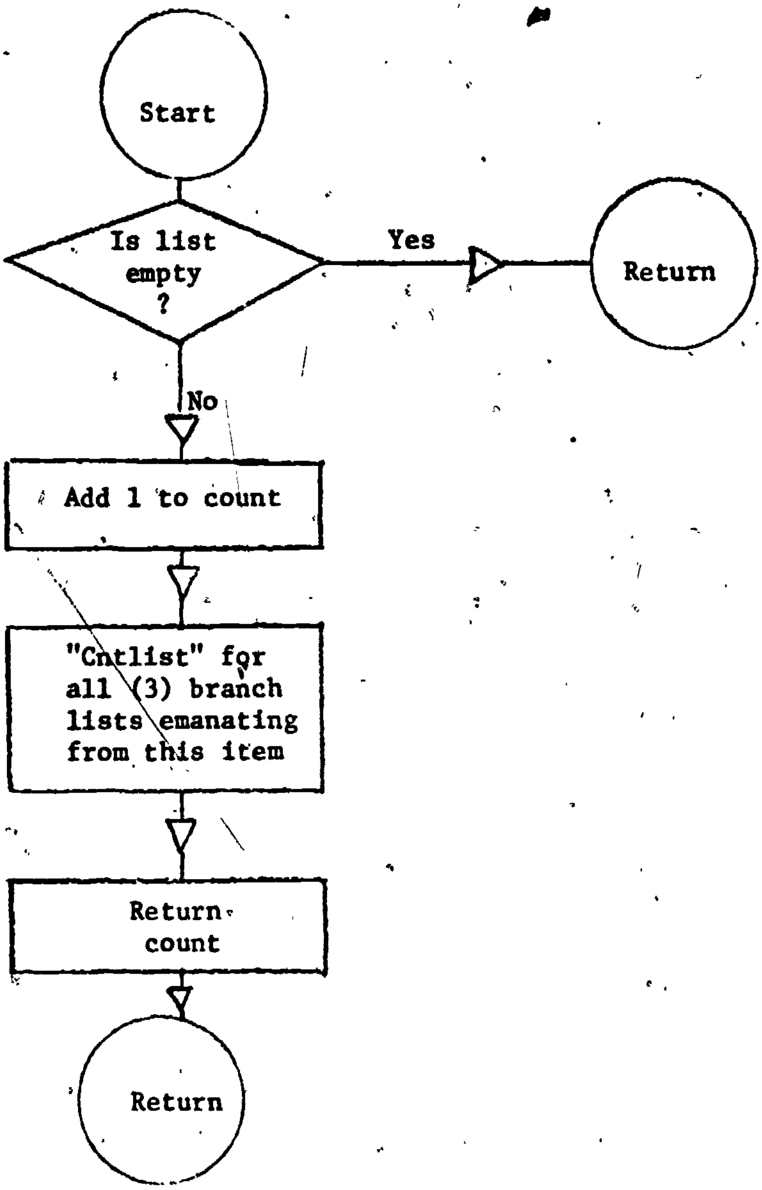


Fig. 47

PRINTSET:
PRINT THE ELEMENTS OF A SET

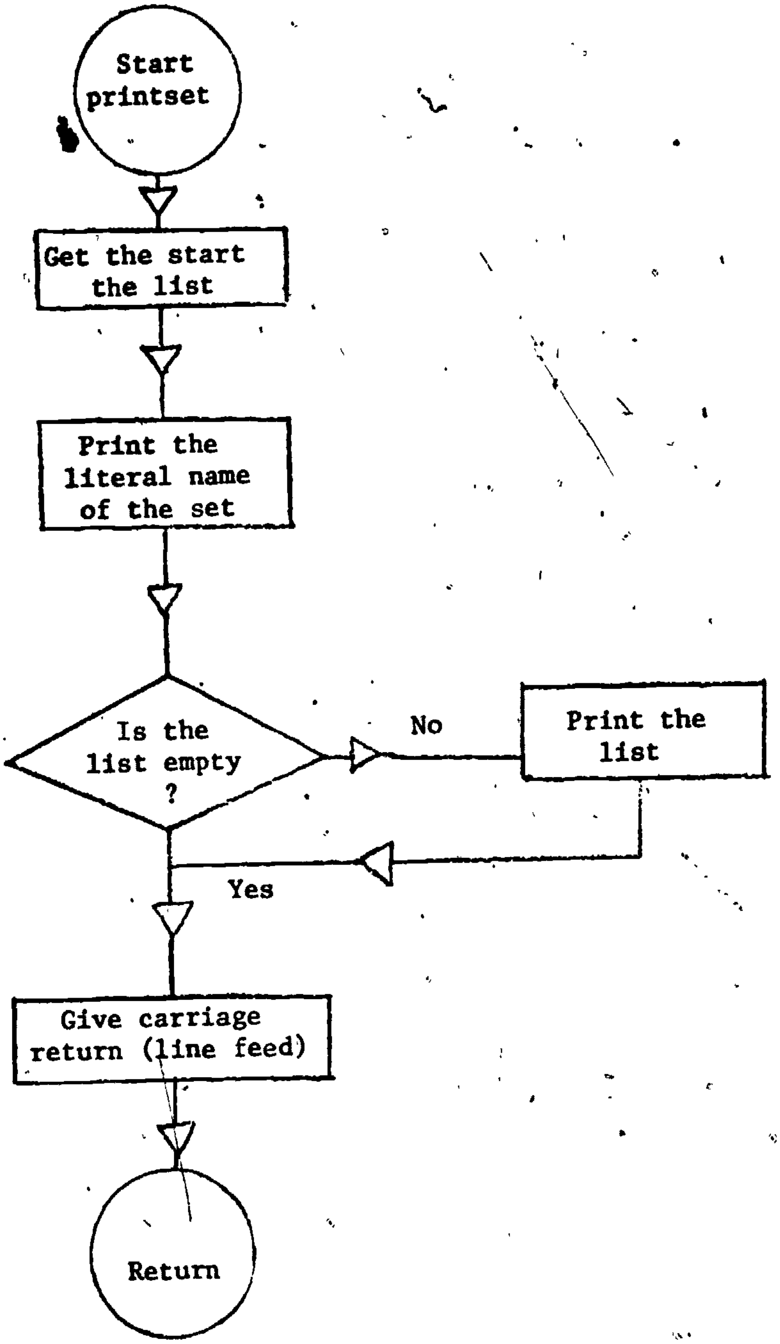


Fig. 48

PRNLIST: PRINTS A LIST

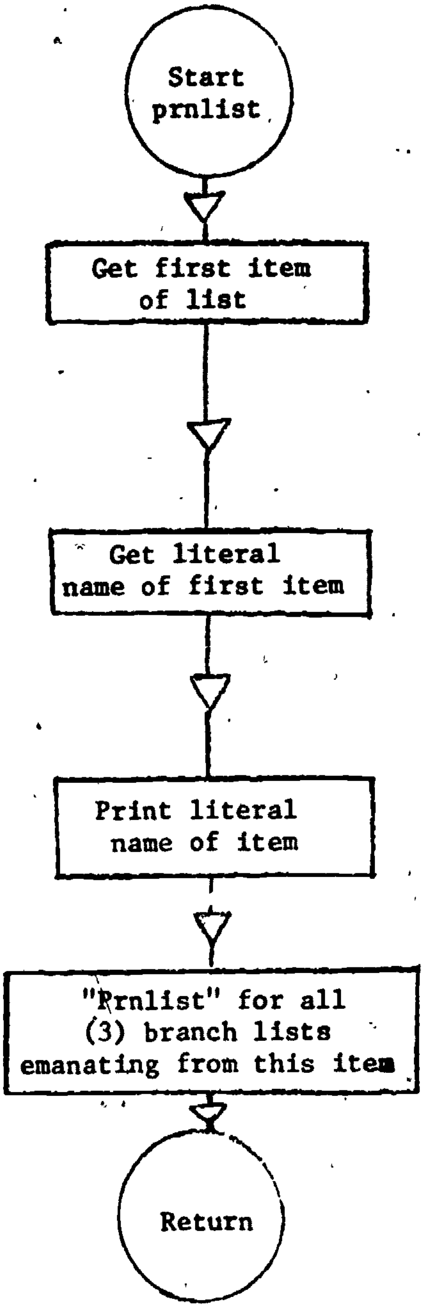


Fig. 49

GARBCAL: COLLECTS NO LONGER NEEDED LISTS AND PUTS LINKS ON FREE LIST

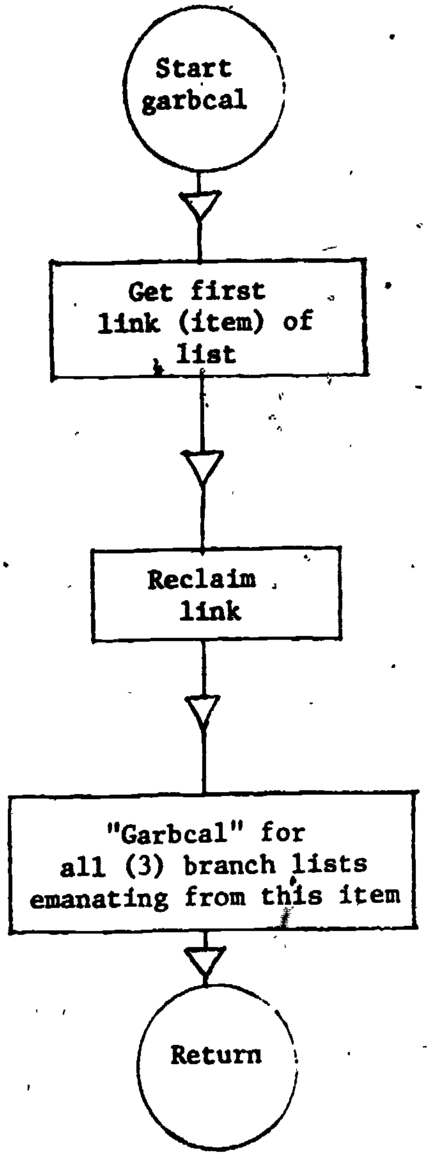


Fig. 50

MAKENUL:
MAKE THIS SET NULL

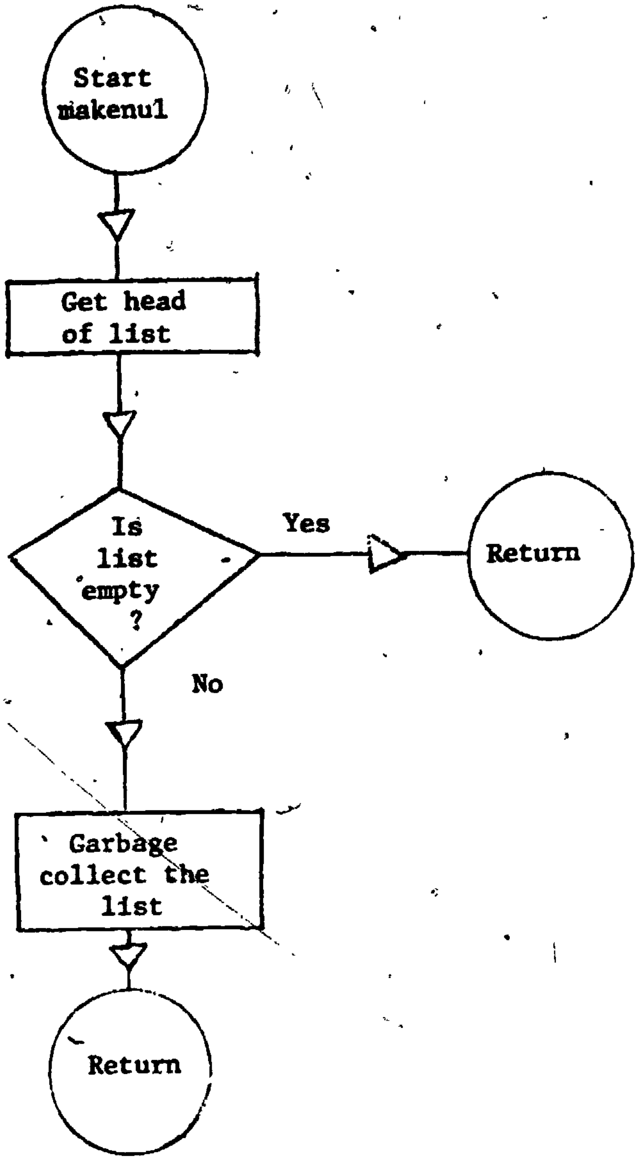


Fig. 51

DIFSET:
GIVES DIFFERENCE OF TWO SETS

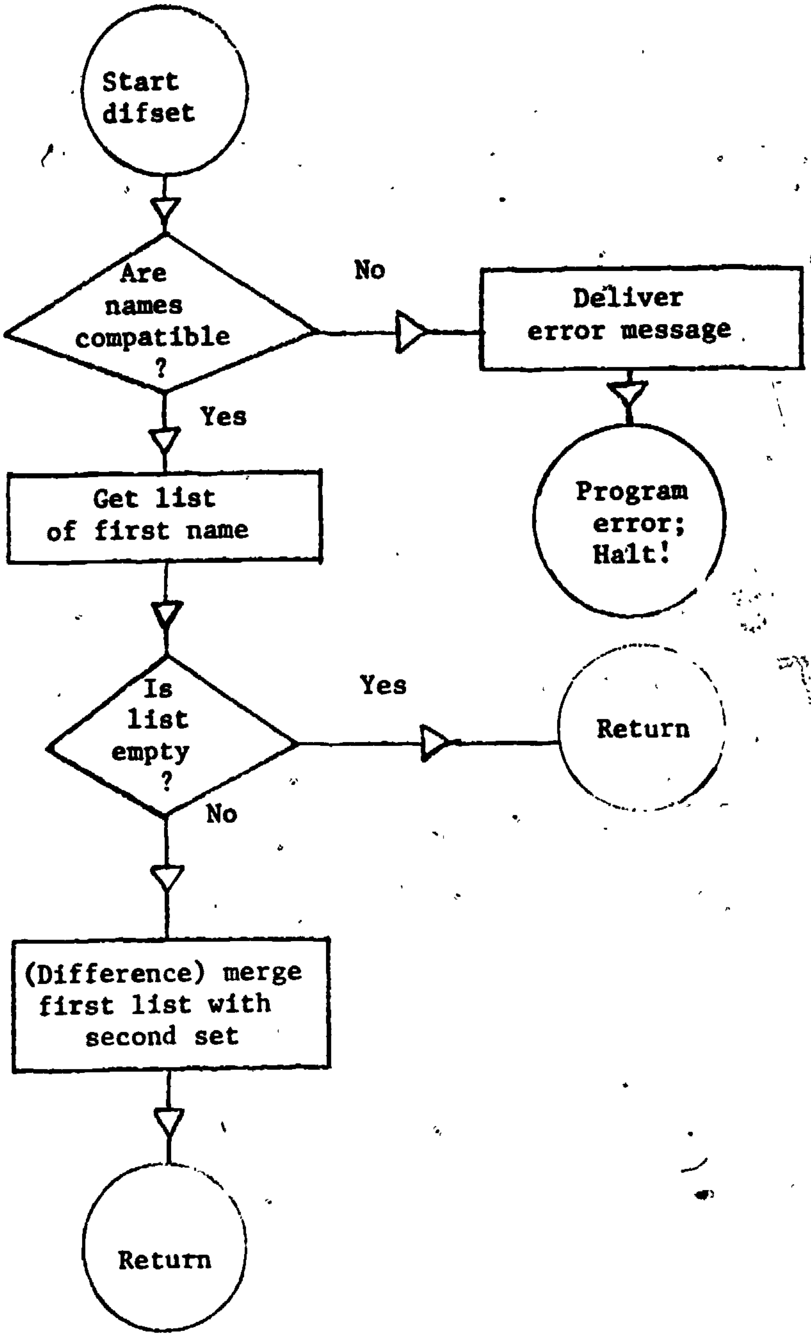


Fig. 52

PRODSET:
GIVES INTERSECTION OF TWO SETS

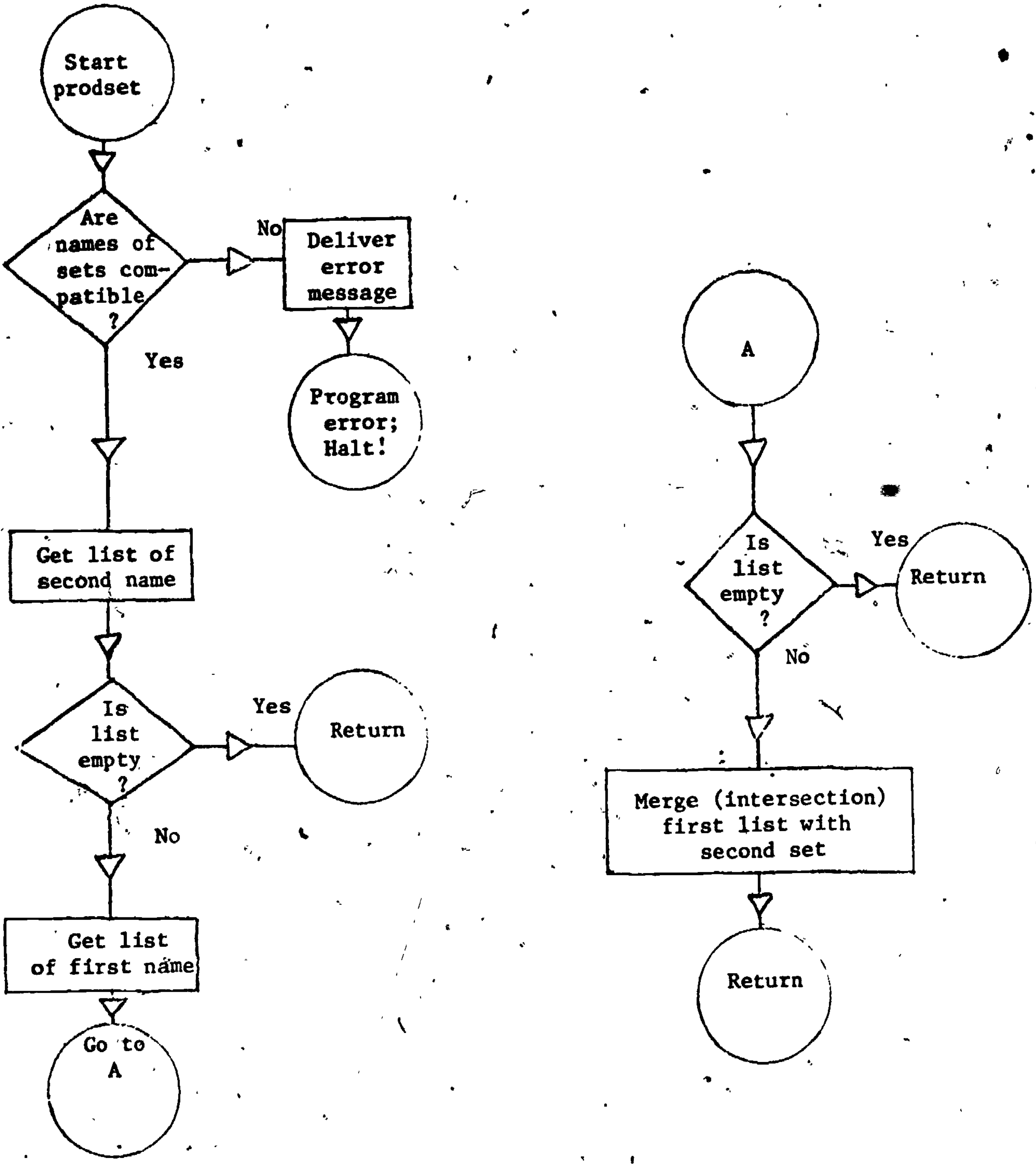


Fig. 53

SUBSET:
CHECKS IF ONE SET IS A SUBSET OF ANOTHER

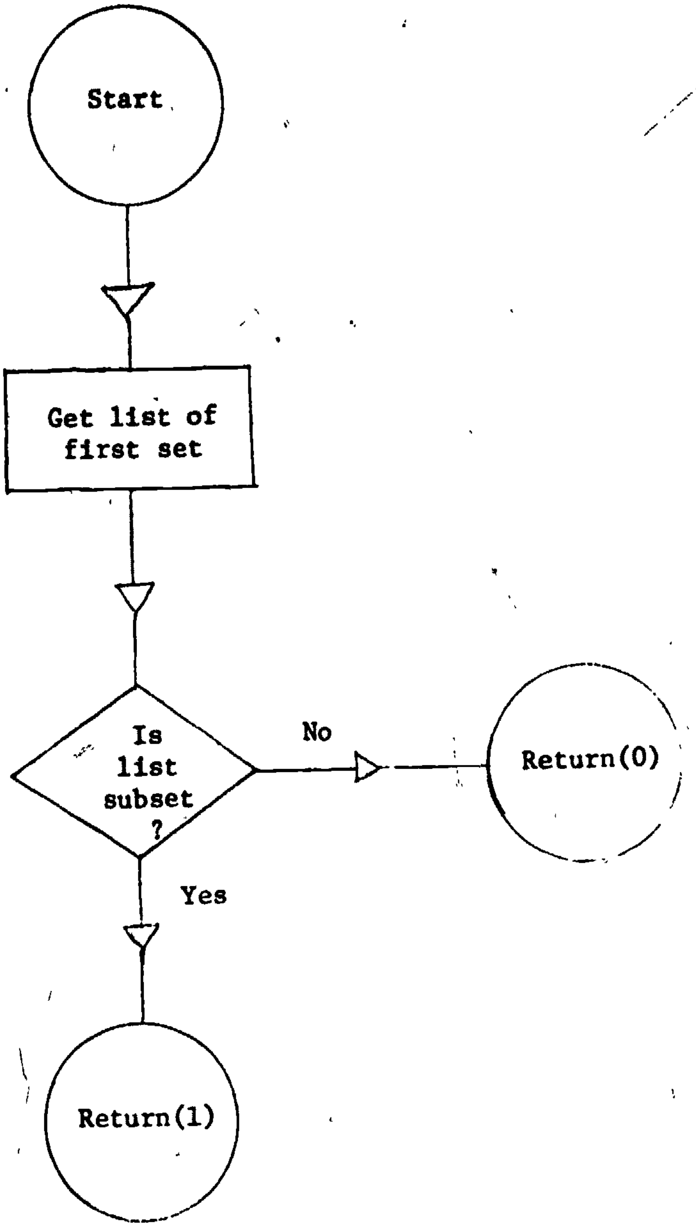


Fig. 54

SETEQ:
CHECKS FOR IDENTITY BETWEEN SETS

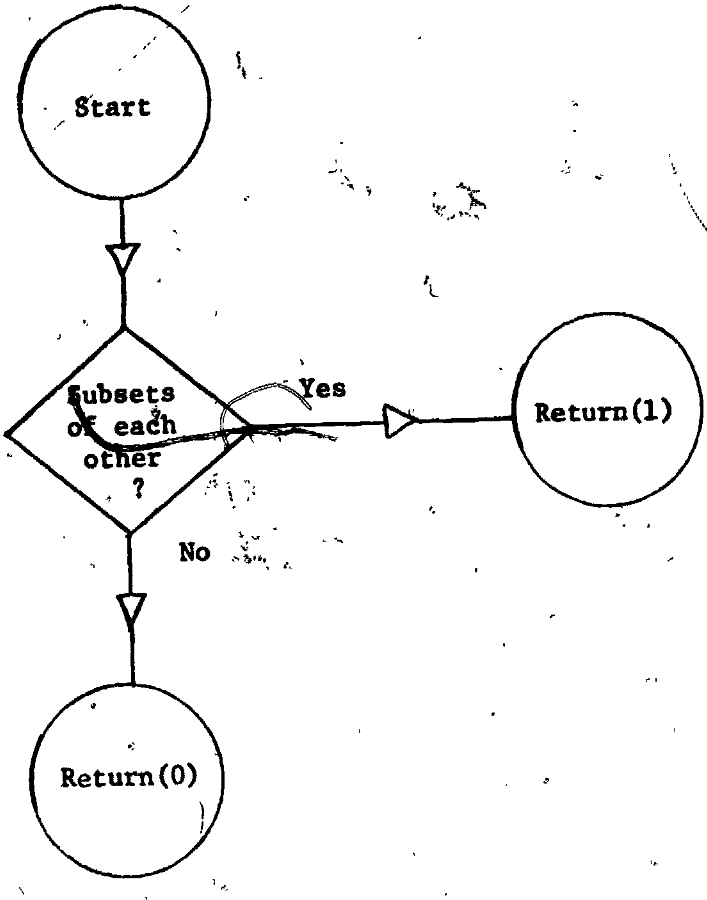


Fig. 55

SUBLIST:
CHECK IF LIST IS A SUBLIST OF A NAMED LIST

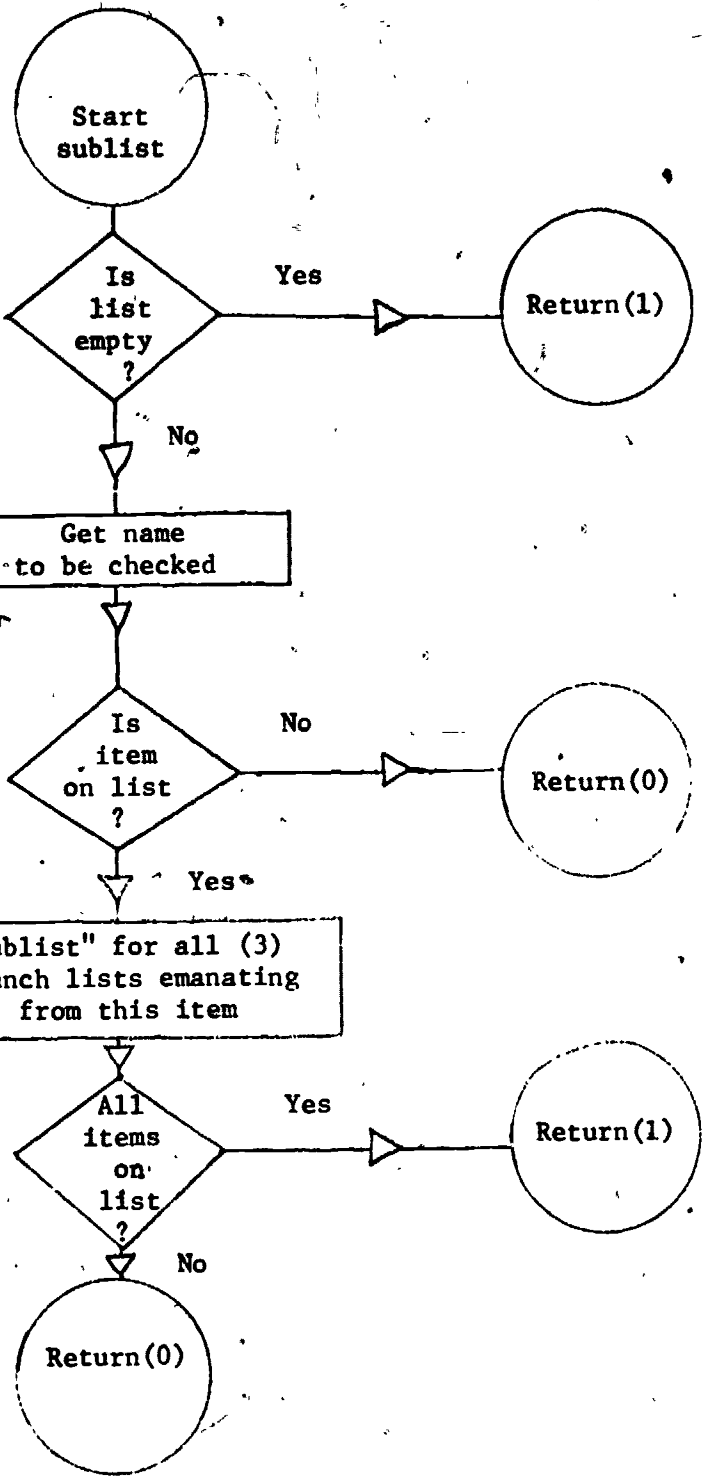


Fig. 56

ADDSET:
GIVE UNION OF TWO NAMED SETS

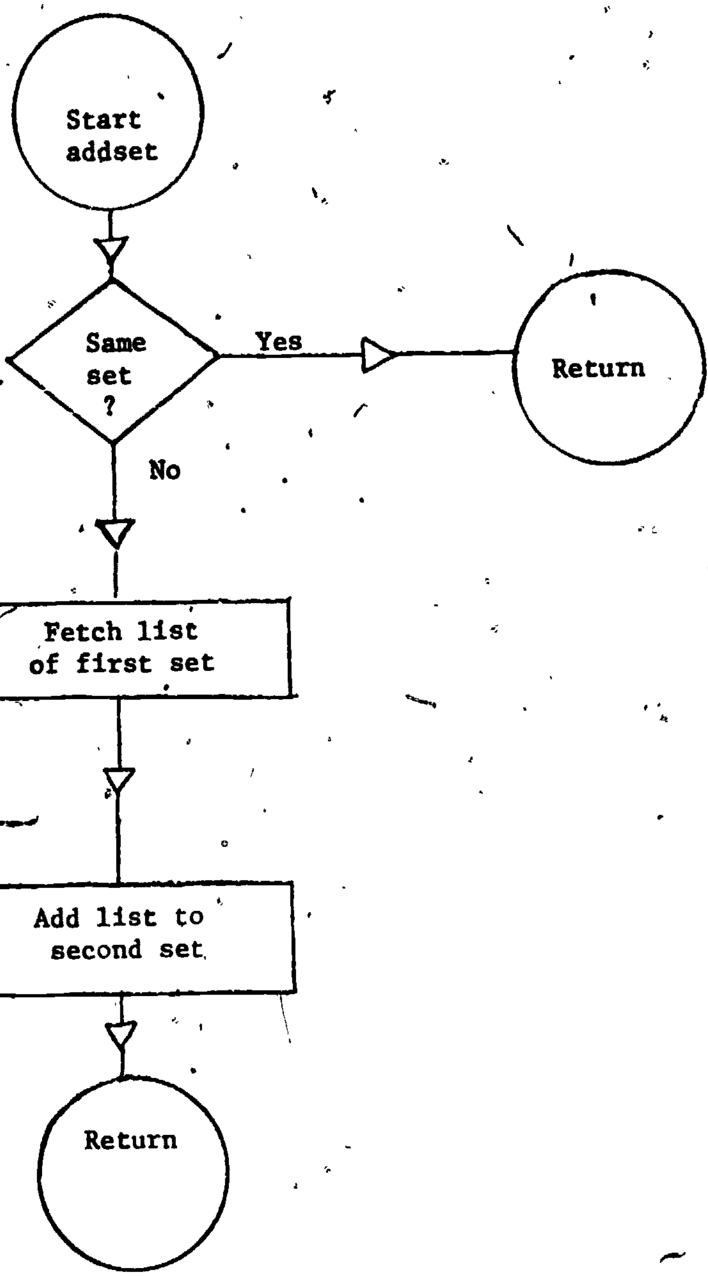


Fig. 57

PDLIST:

PUTS ITEM FROM (A) LIST ON OTHER NAMED (C) LIST DEPENDING ON ANOTHER NAMED (B) LIST

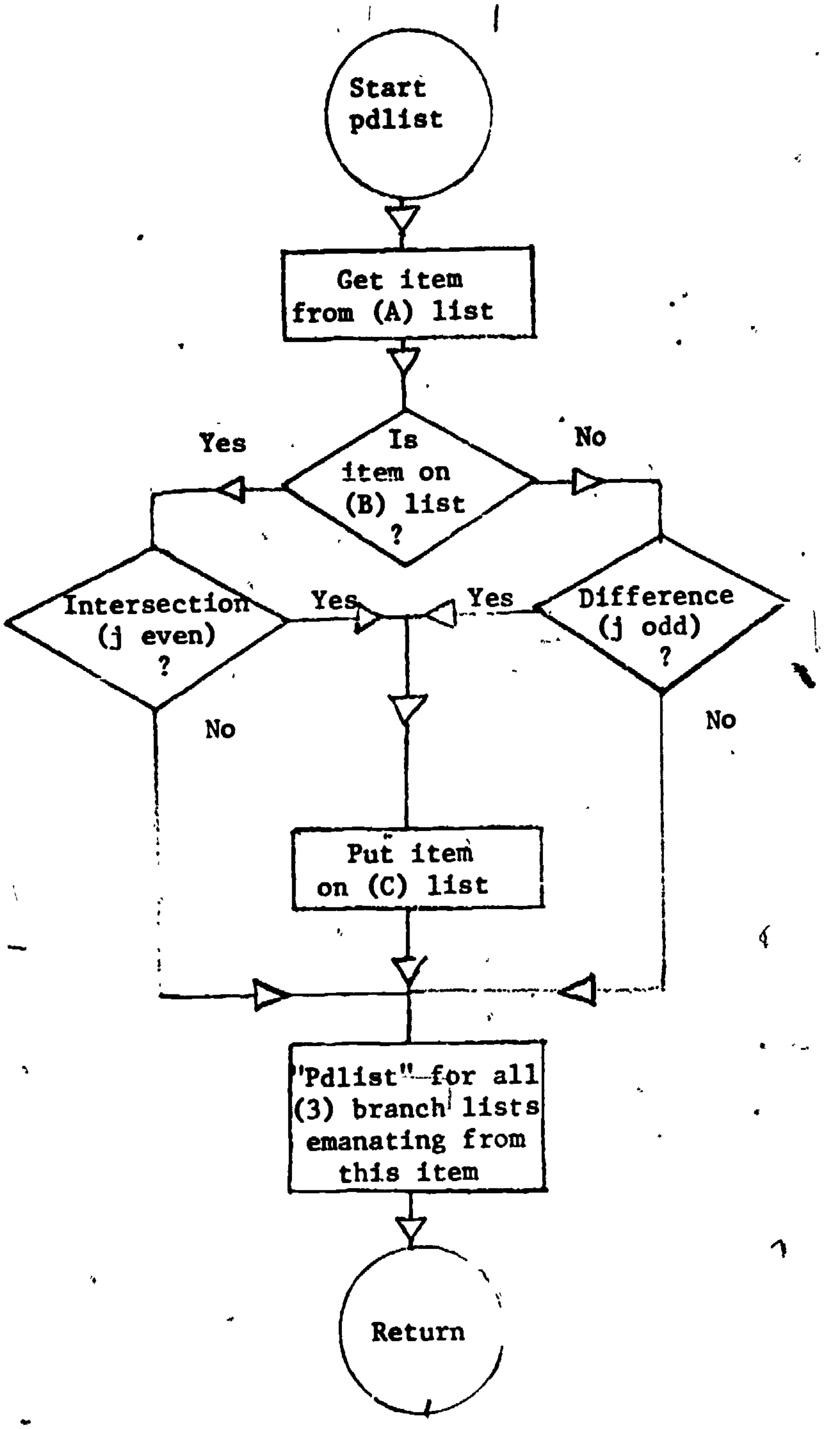


Fig. 58

ONLIST:
CHECKS TO SEE IF ITEM IS ON NAMED LIST

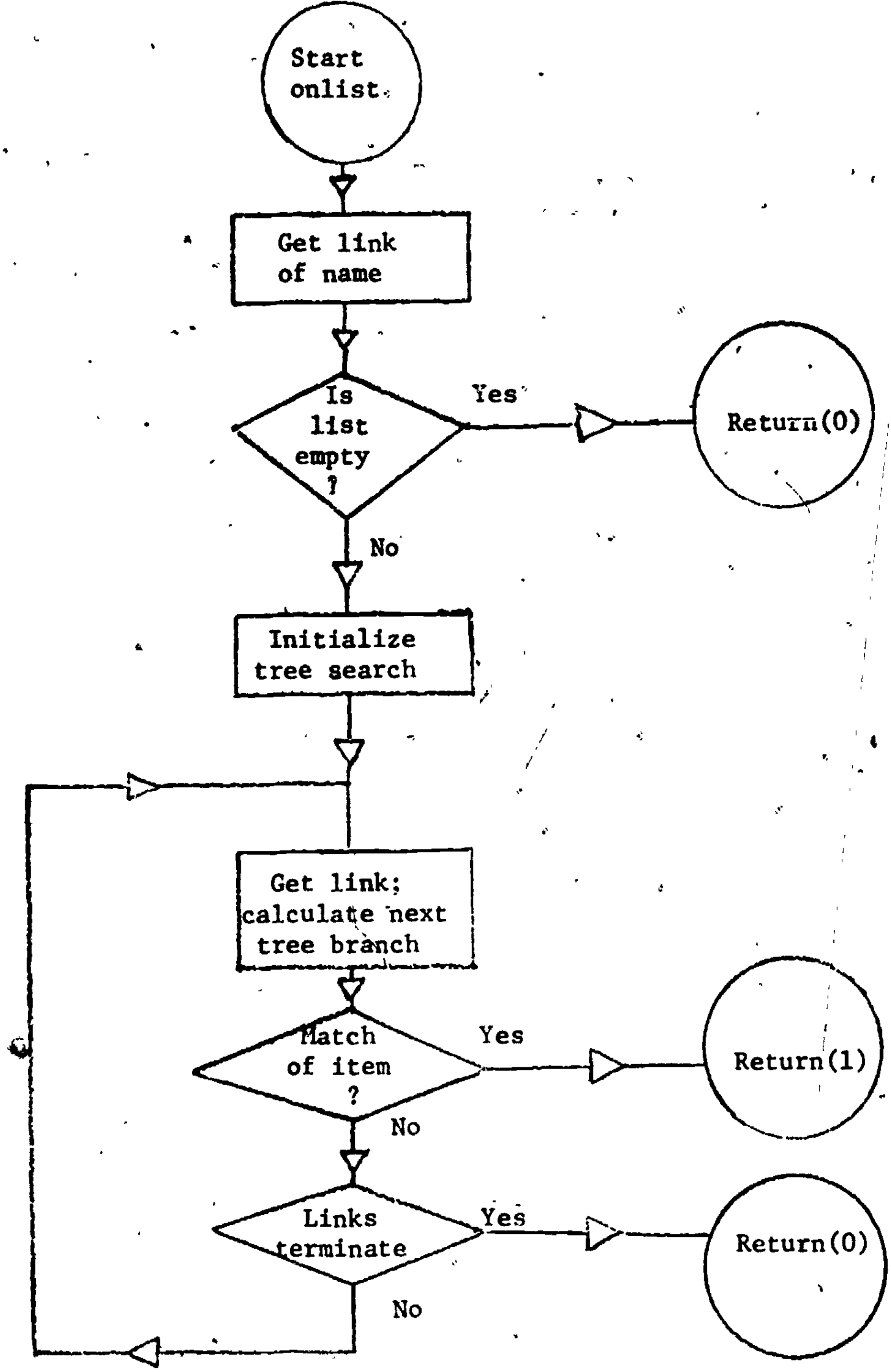


Fig. 59

PUTITEM:
PLACES AN ITEM ON NAMED LIST

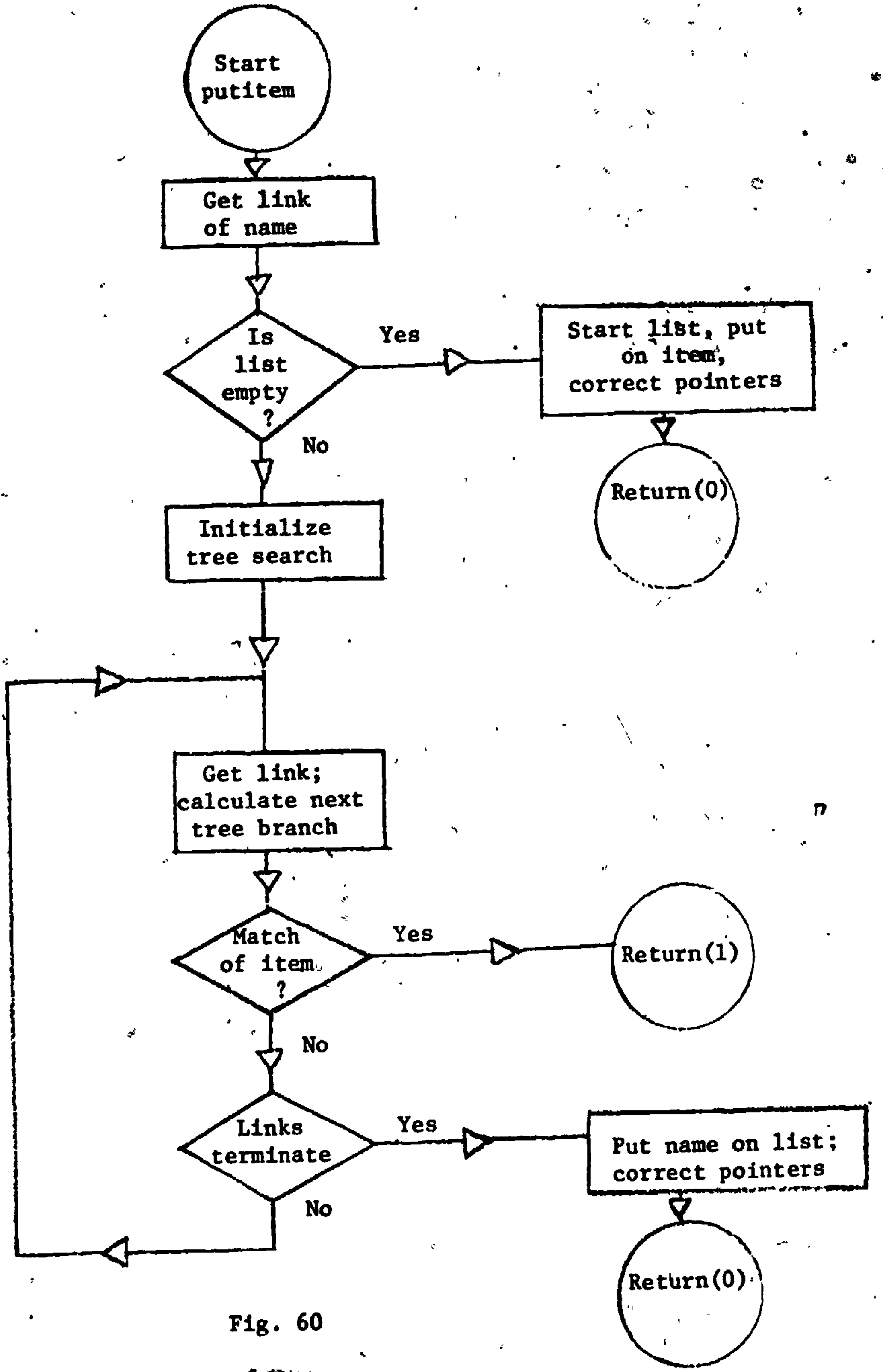
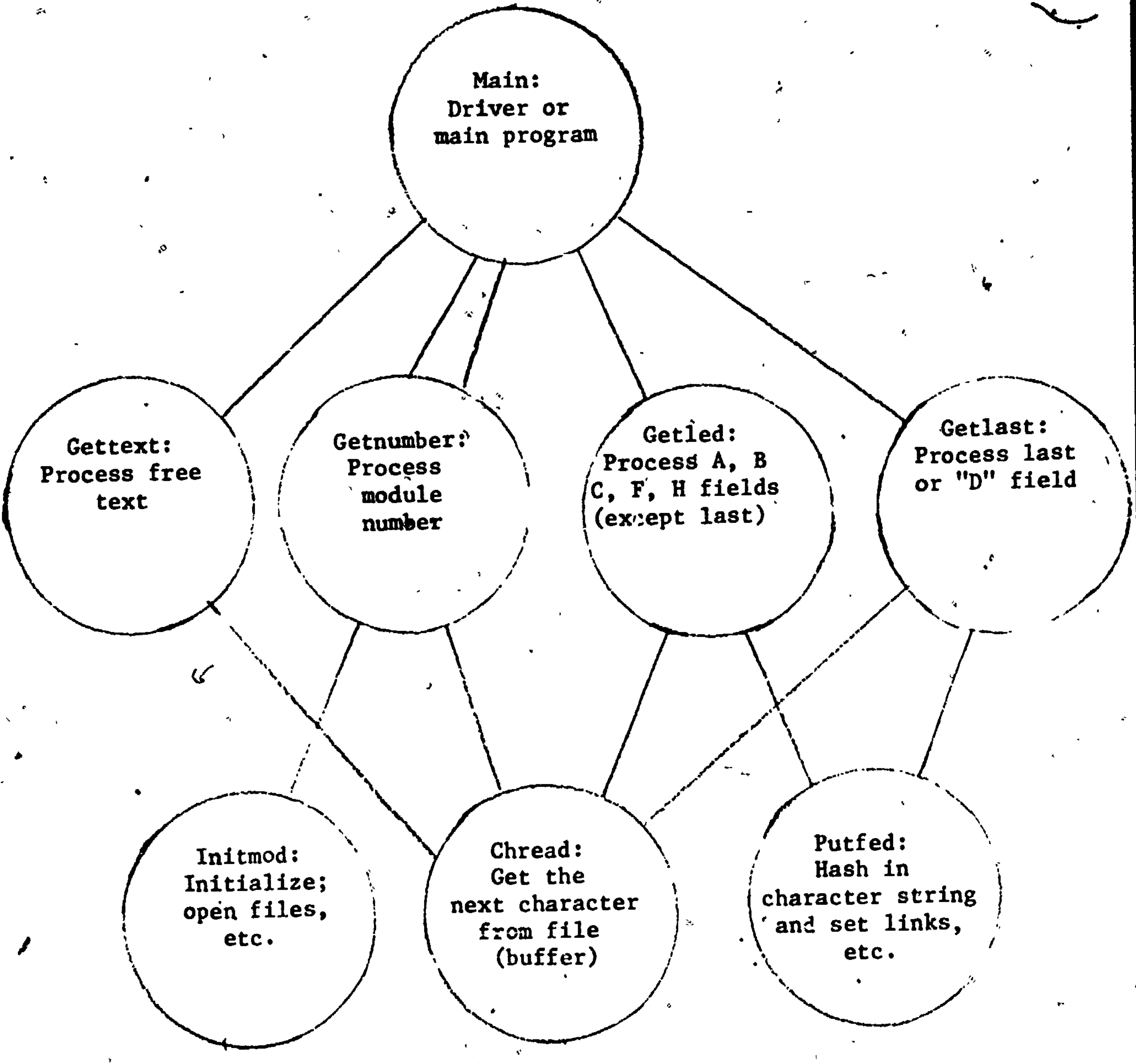


Fig. 60

LATTICE DIAGRAM
MODULE PROCESSING
BLOCK* (mp.c)



*Procedures called in lower blocks not shown.

Fig. 61

PUTFLD:
HASHES IN CHARACTER STRINGS AND SETS LINKS

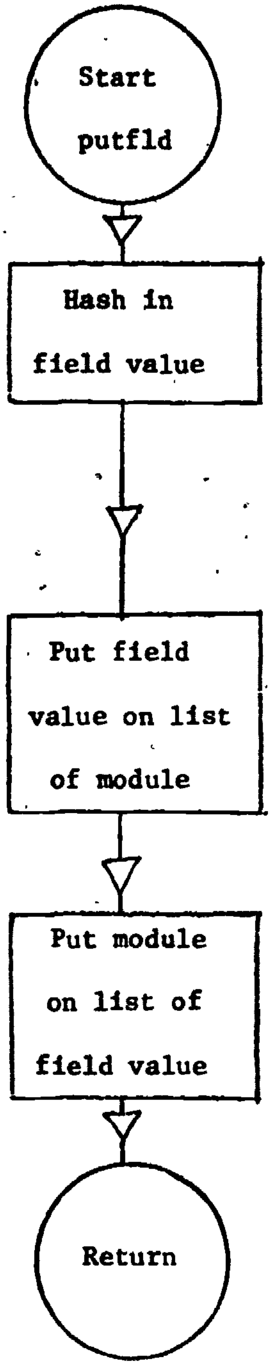


Fig. 62

CHREAD: READS CHARACTER FROM FILE (BUFFER)

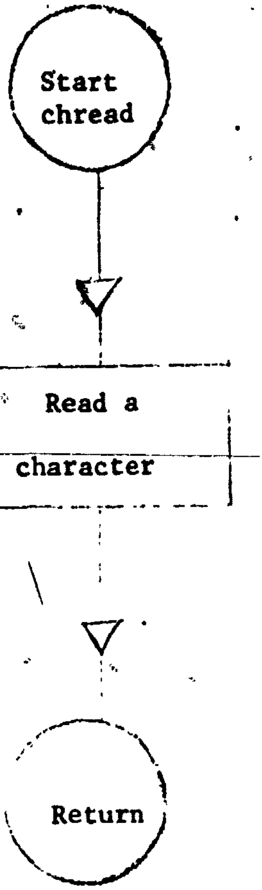


Fig. 63

INITMOD:
INITIALIZES MODULE FILE AND COUNTERS

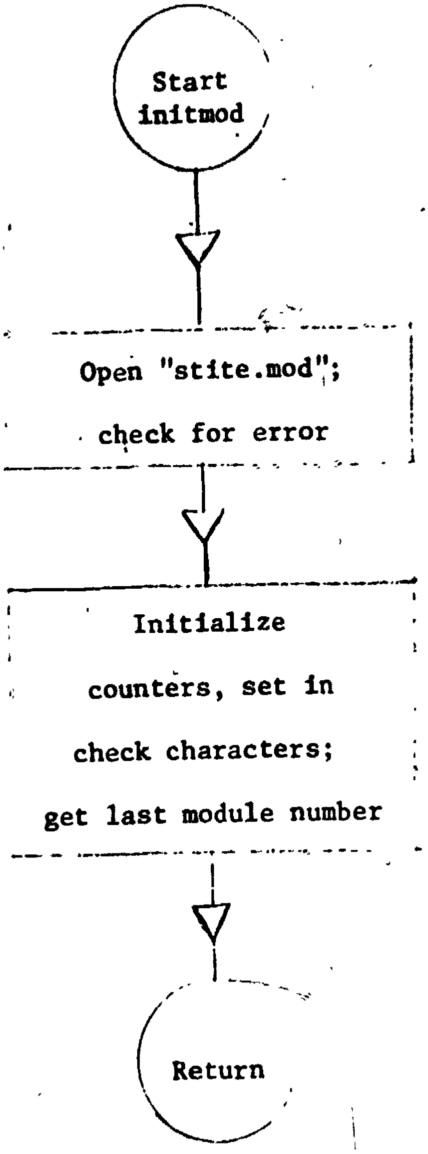


Fig. 64

GETMODN:
PROCESS MODULE NAME (NUMBER)

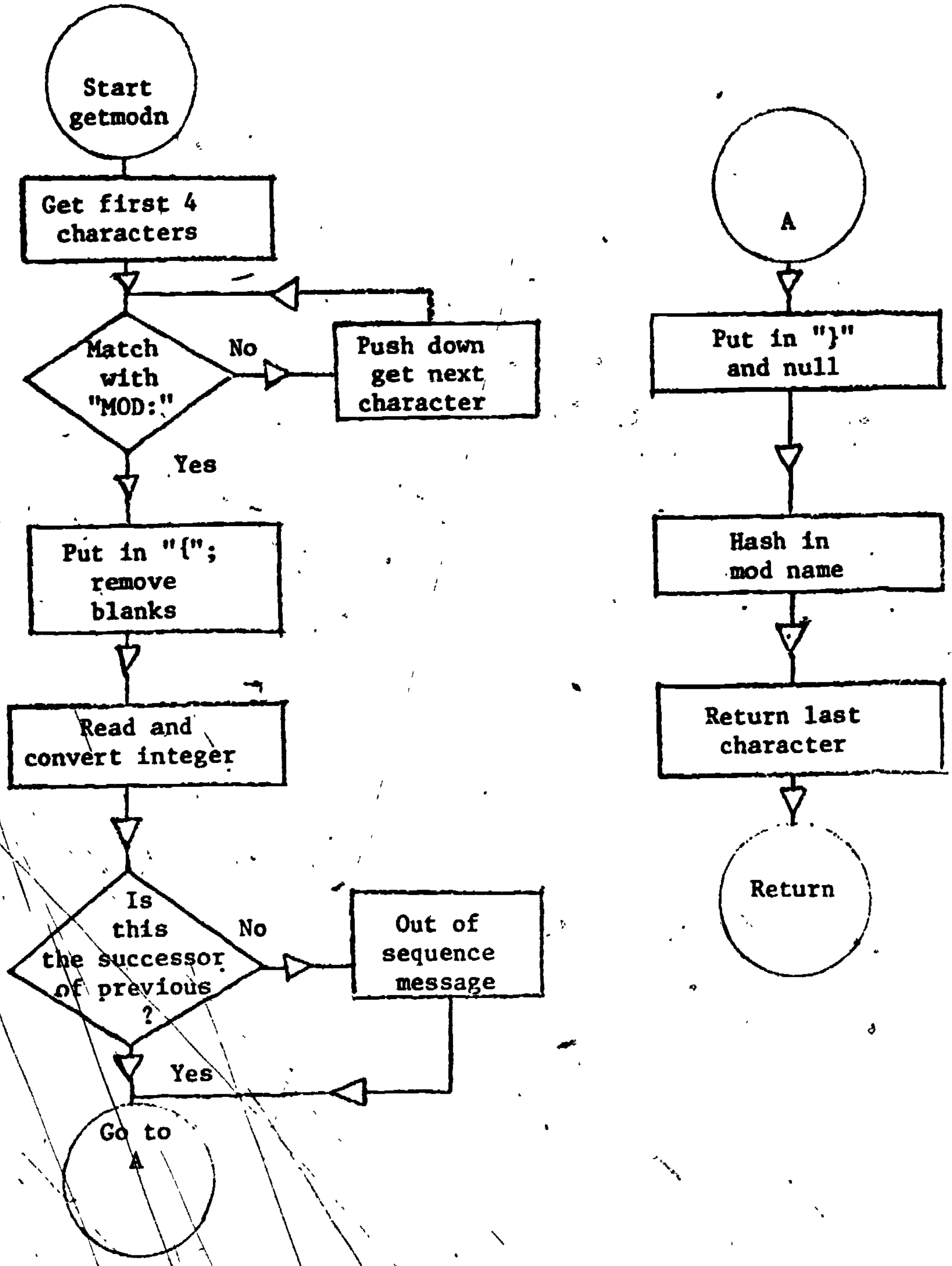


Fig. 65

GETTEXT:
PROCESS FREE TEXT

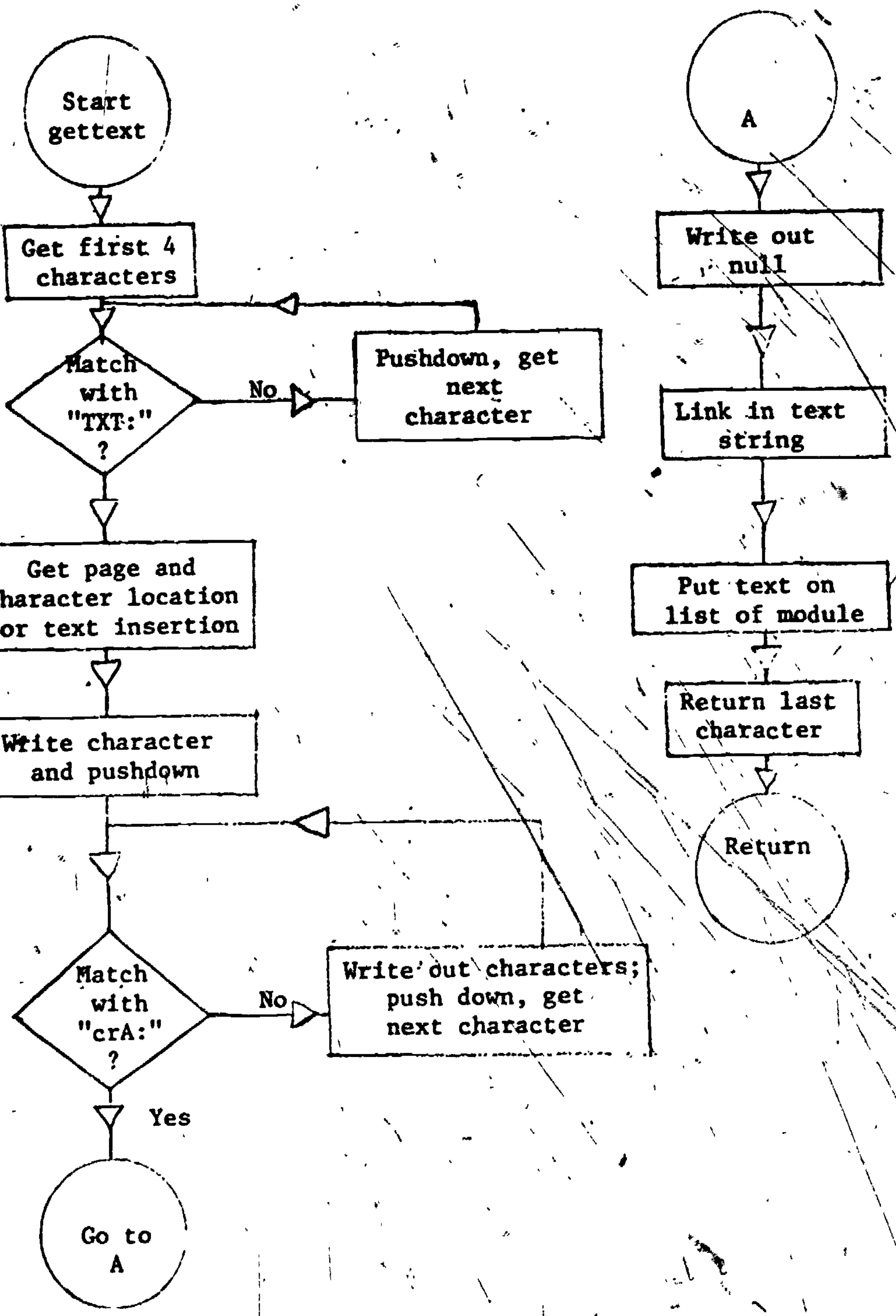


Fig. 66

GETFLD:
PROCESS A, B, C, F, H FIELDS
(BUT NOT D)

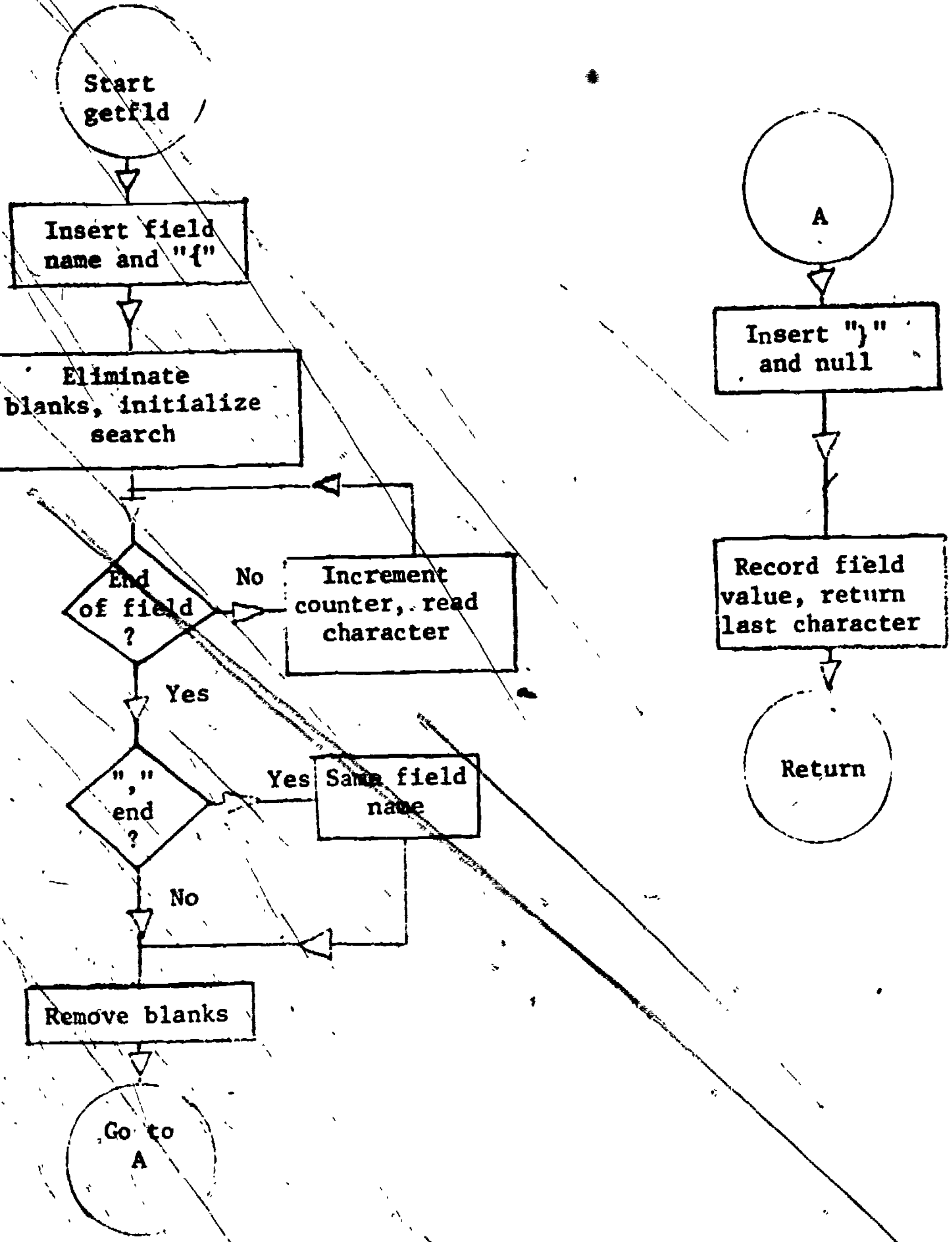


Fig. 67

GETLAST:
PROCESS LAST OR "D" FIELD

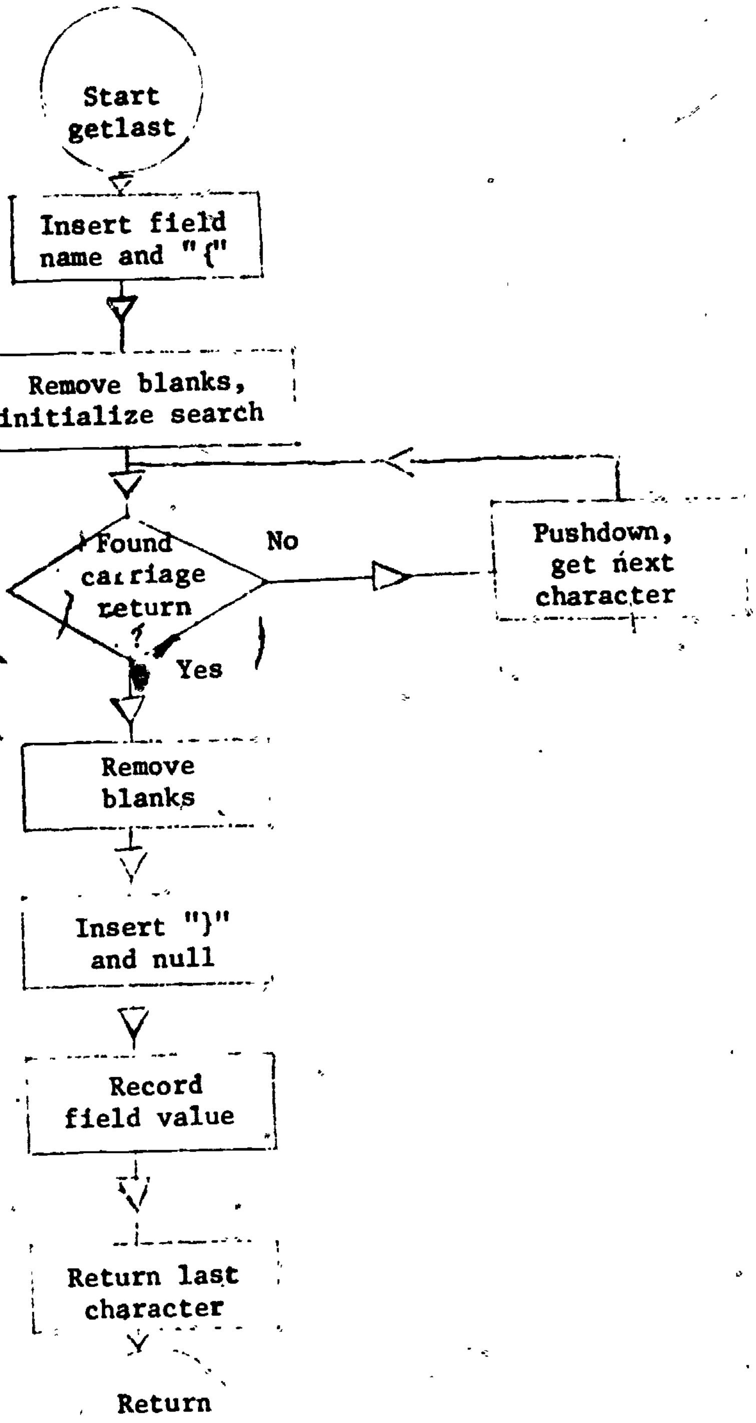


Fig. 68

MAIN

DRIVER OR MAIN PROGRAM FOR MODULE PROCESSING

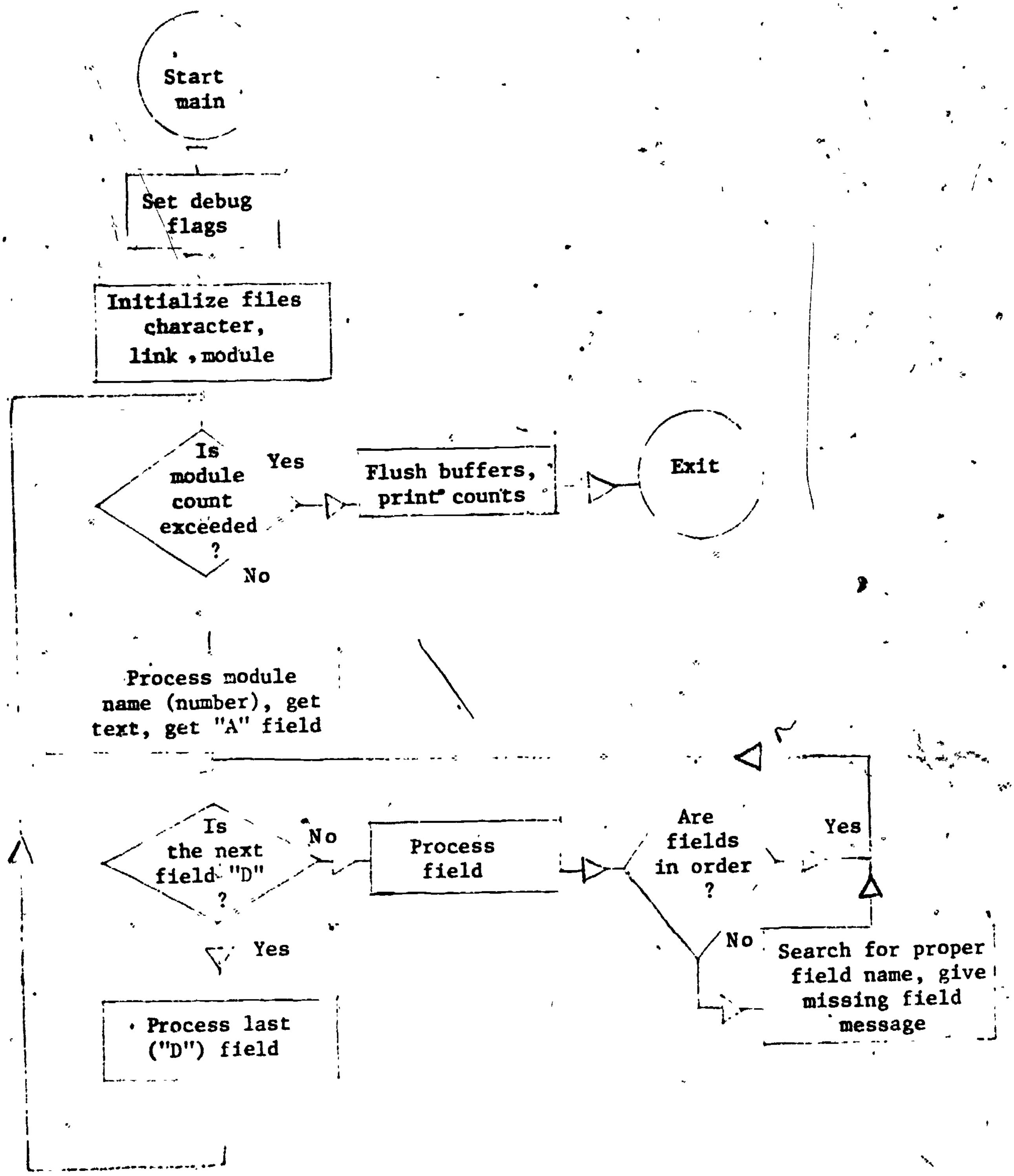


Fig. 69

THE INTERACTIVE (SET MANIPULATION) PROGRAM

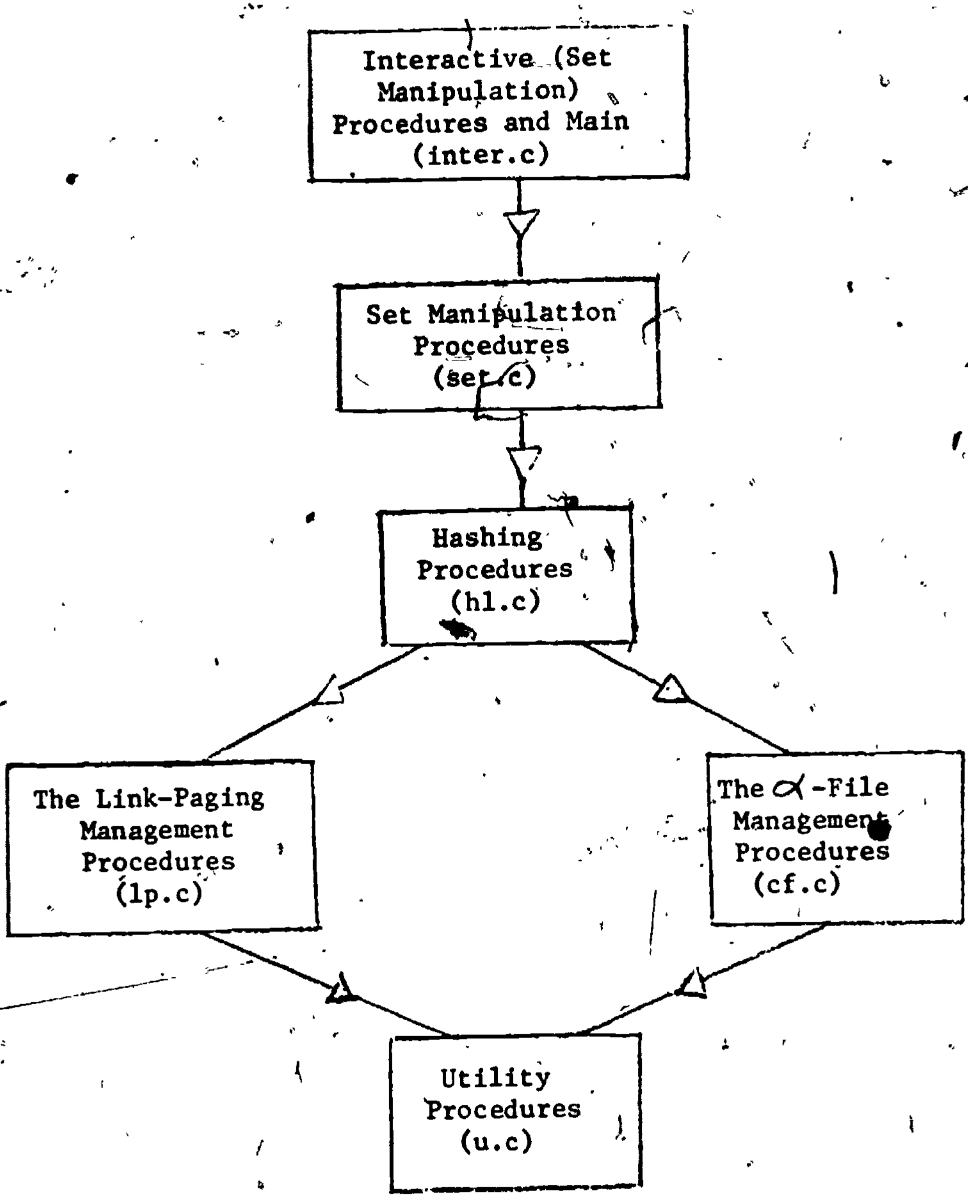


Fig. 70

LATTICE DIAGRAM FOR INTERACTIVE SET MANIPULATION (INTER. C)

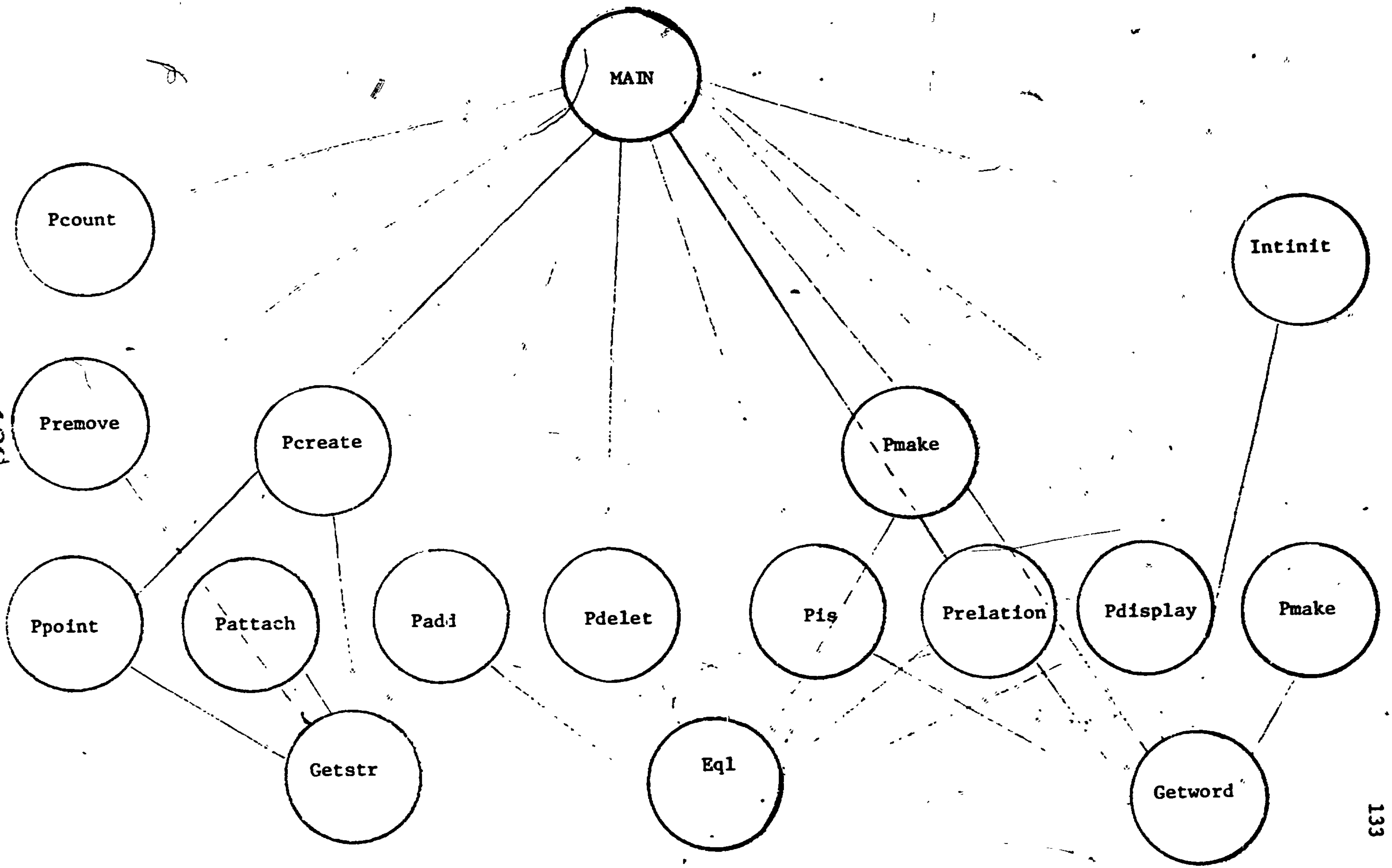


Fig. 71

MAIN
MAIN PROGRAM
OF INTERACTIVE SET MANIPULATOR

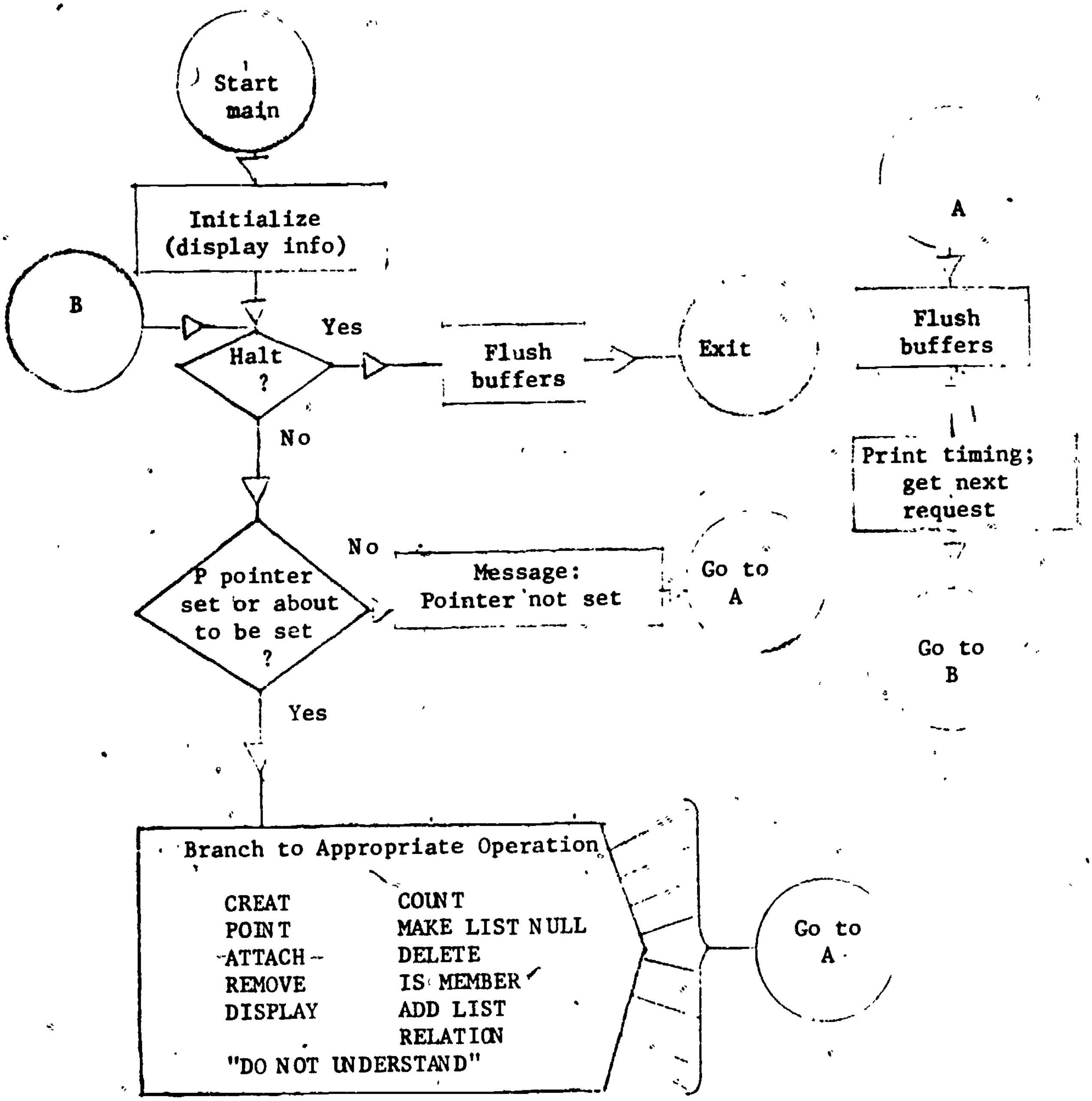


Fig. 72

PMAKE: SET LIST TO NULL

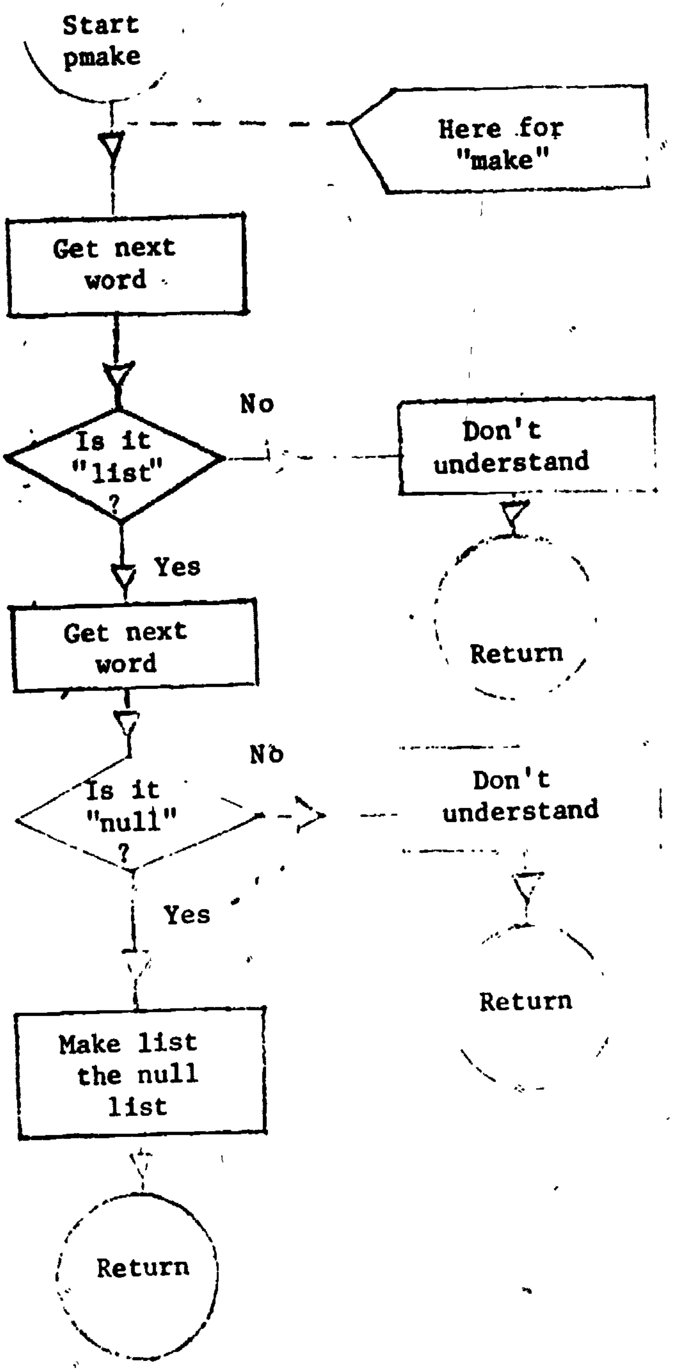


Fig. 73

PREMOVE:
REMOVES SINGLE ITEM FROM LIST

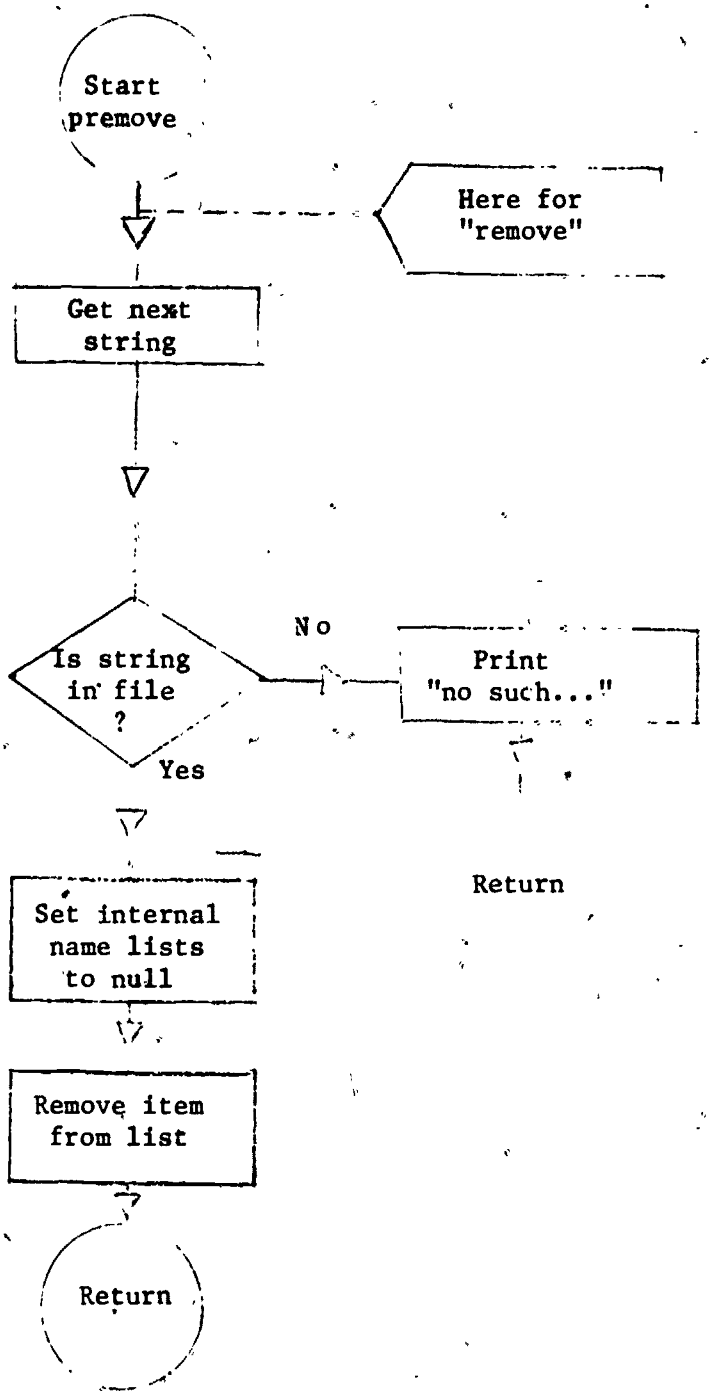


Fig. 74

PDELETE:
REMOVE LIST FROM ANOTHER

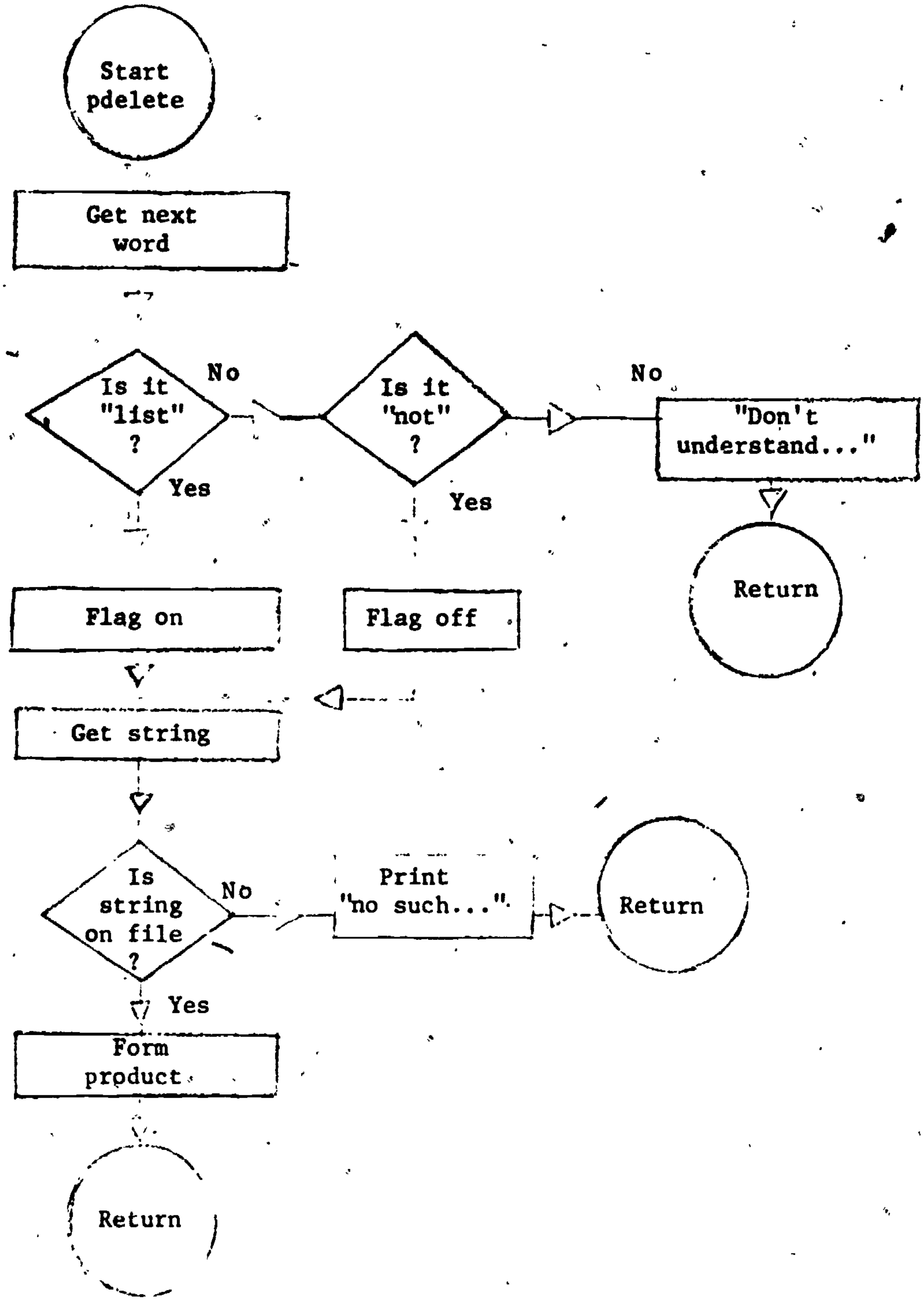


Fig. 75

PADD:
FORMS UNION OF TWO SETS

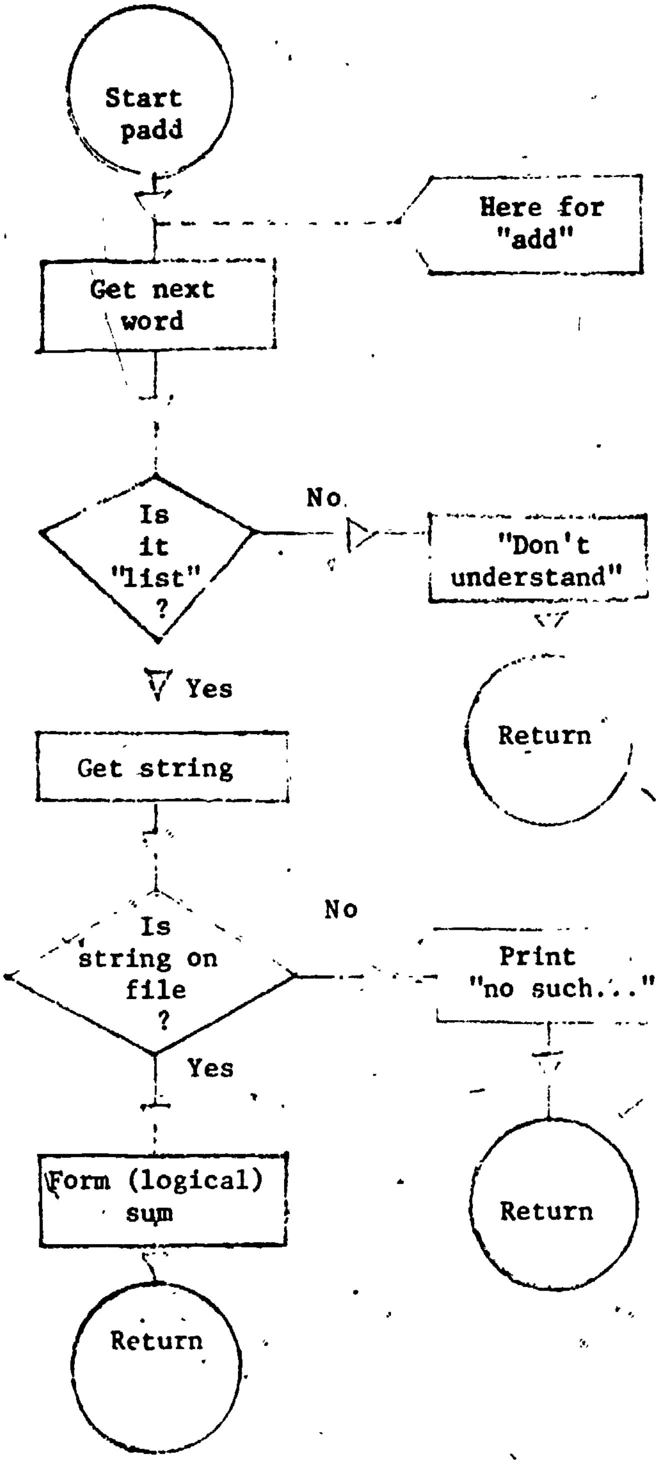


Fig. 76

PRELATION:
DETERMINES RELATION BETWEEN TWO SETS

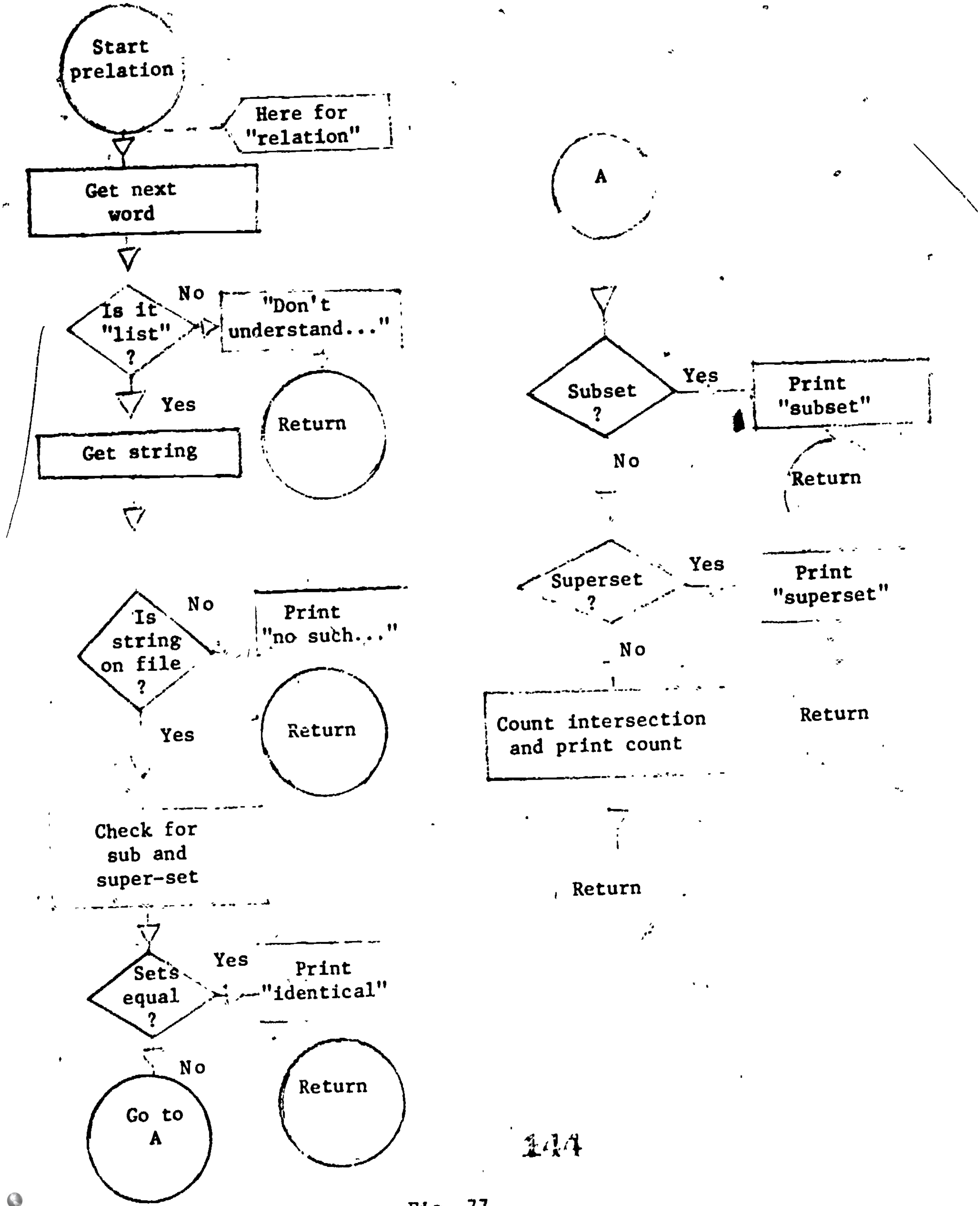


Fig. 77

PIS:
CHECK FOR SET MEMBERSHIP

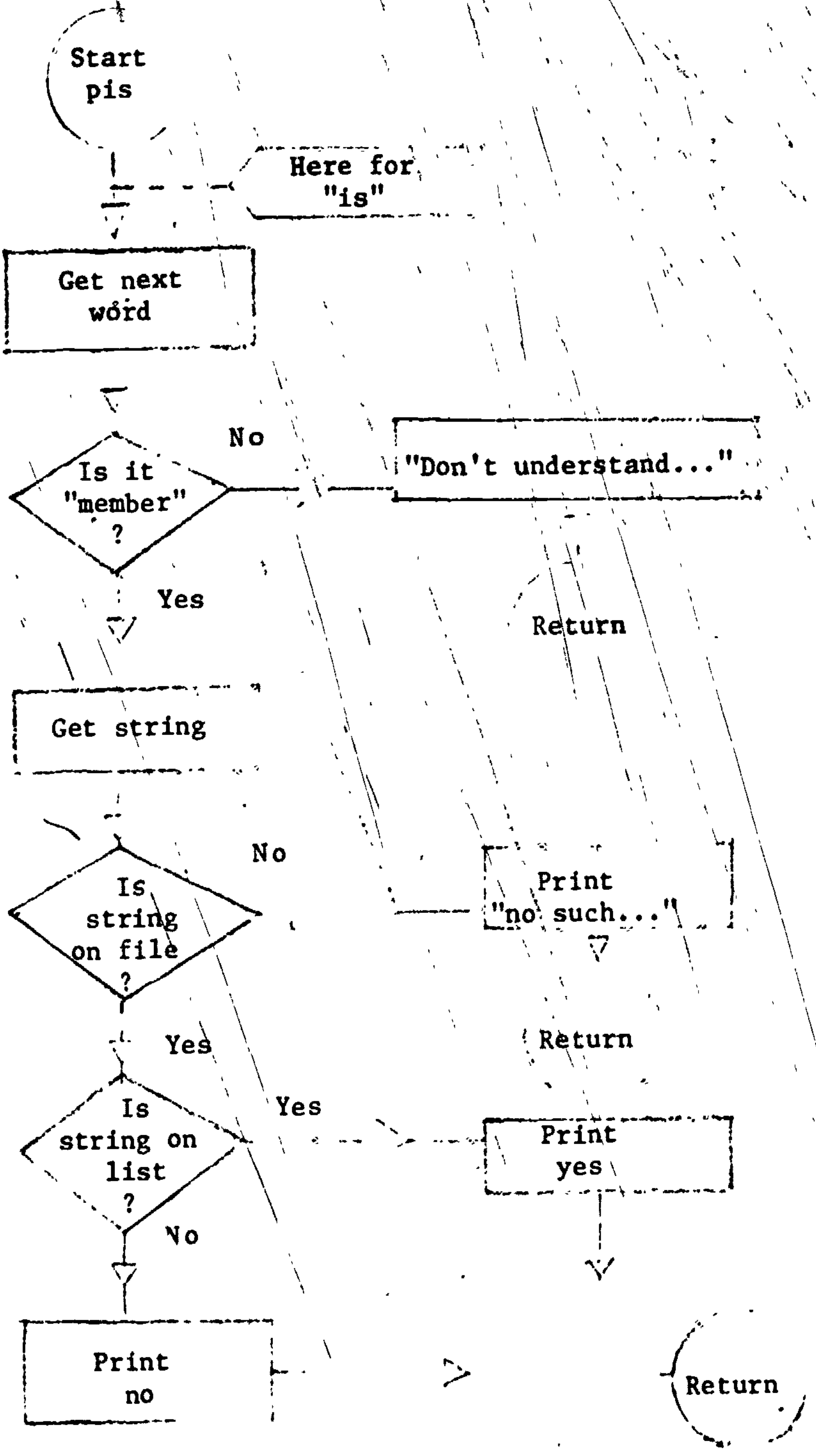


Fig. 78

PATTACH:
PUTS ELEMENT IN SET

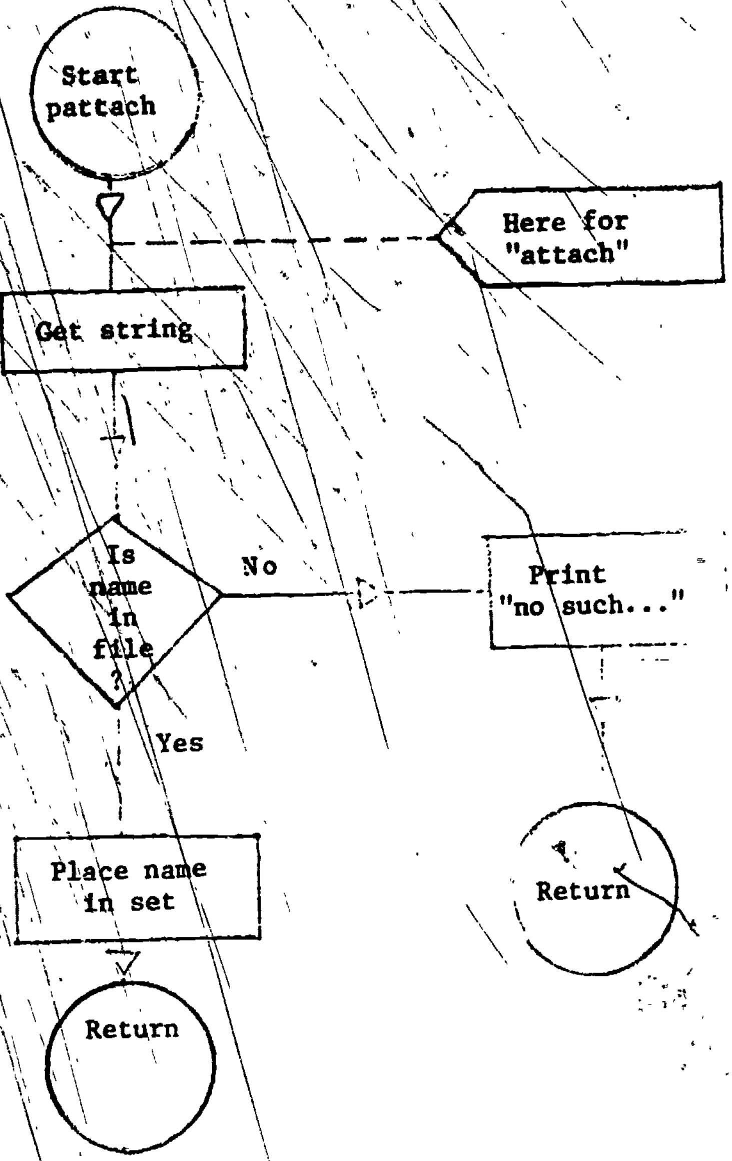


Fig. 79

PCOUNT:
COUNTS THE NUMBER OF ELEMENTS IN A SET

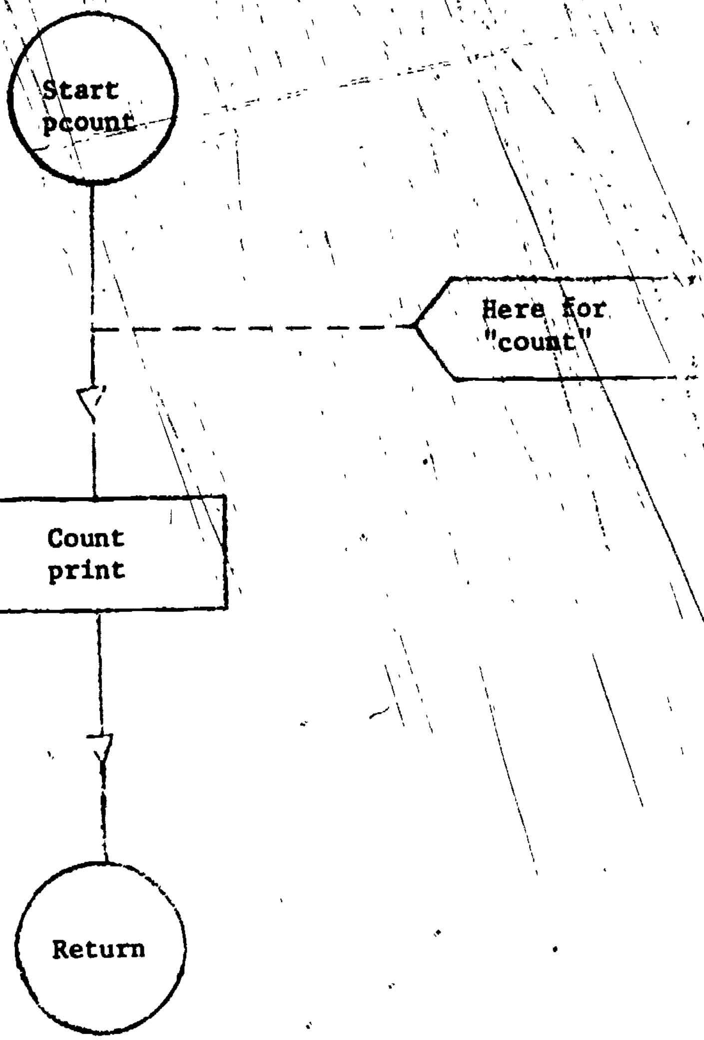


Fig. 80

PDISPLAY:
DISPLAY ITEM OR LIST

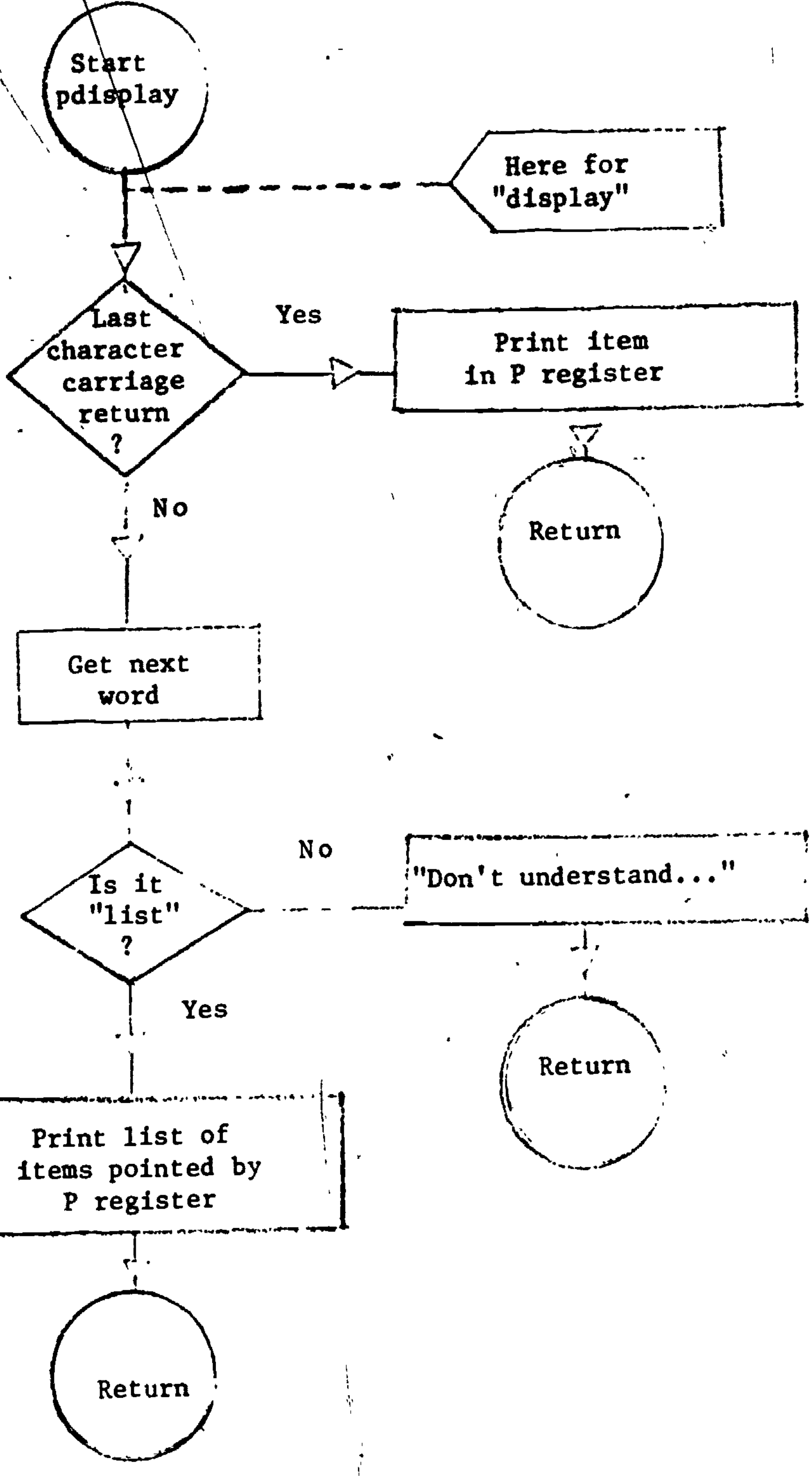


Fig. 81

PPOINT:
SET POINTER TO APPROPRIATE SET

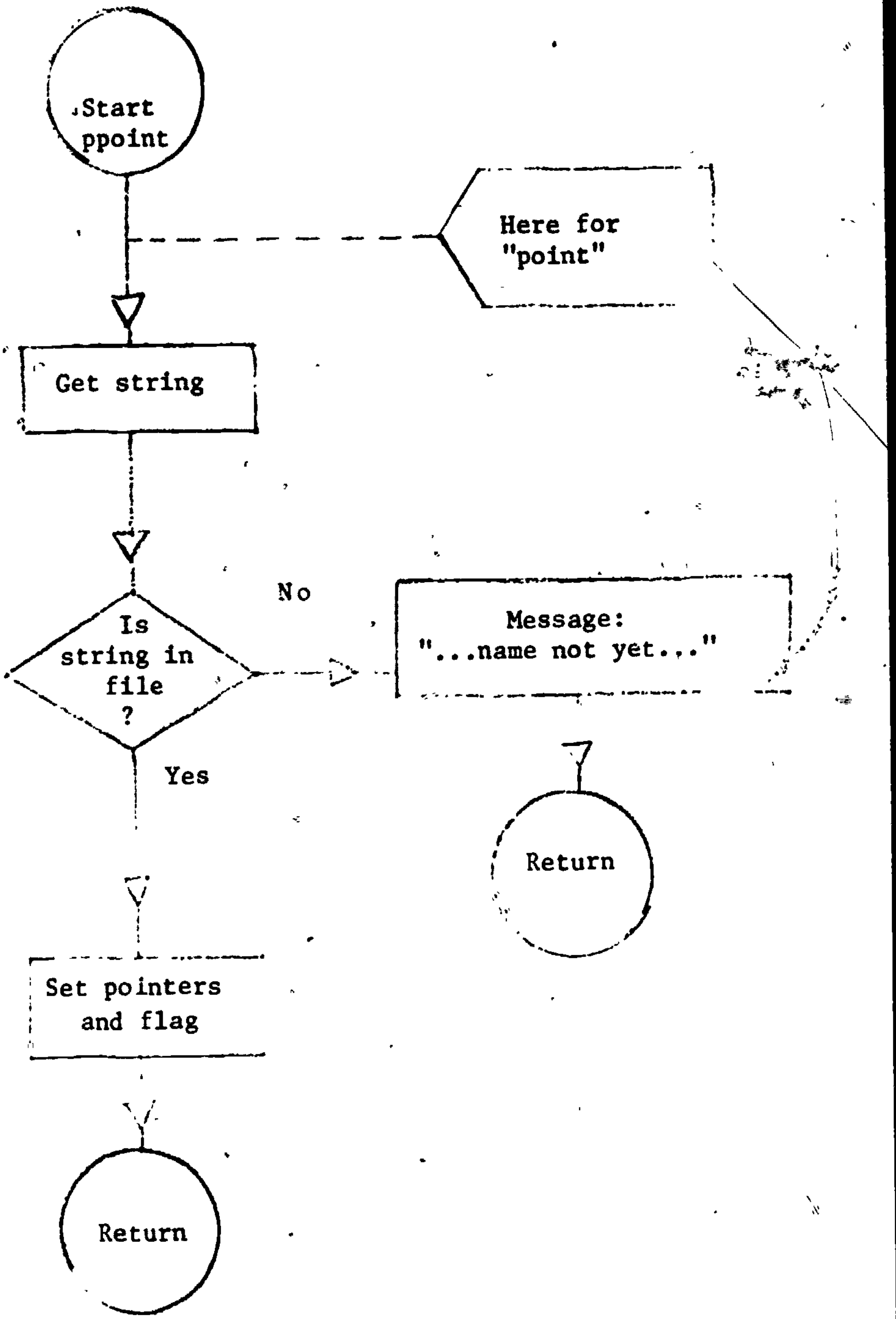


Fig. 82

PCREATE:
INTRODUCES NEW NAME

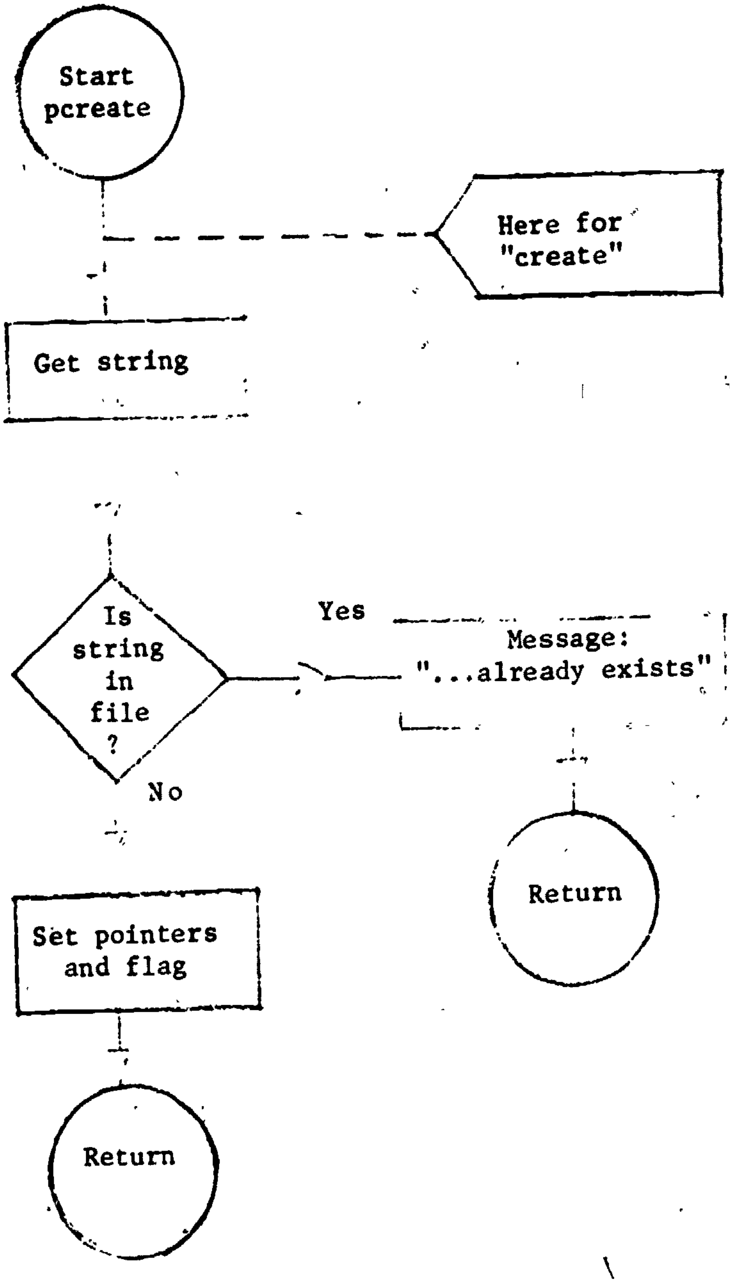


Fig. 83

GETWORD:
GETS THE NEXT WORD

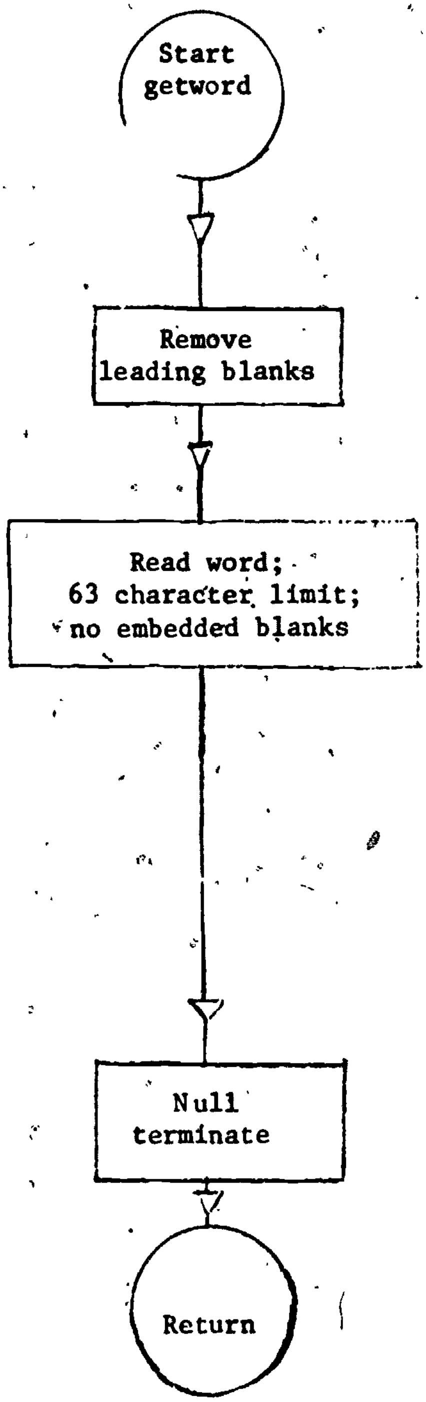


Fig. 84

INTINIT: INITIALIZE INTERACTIVE MODE

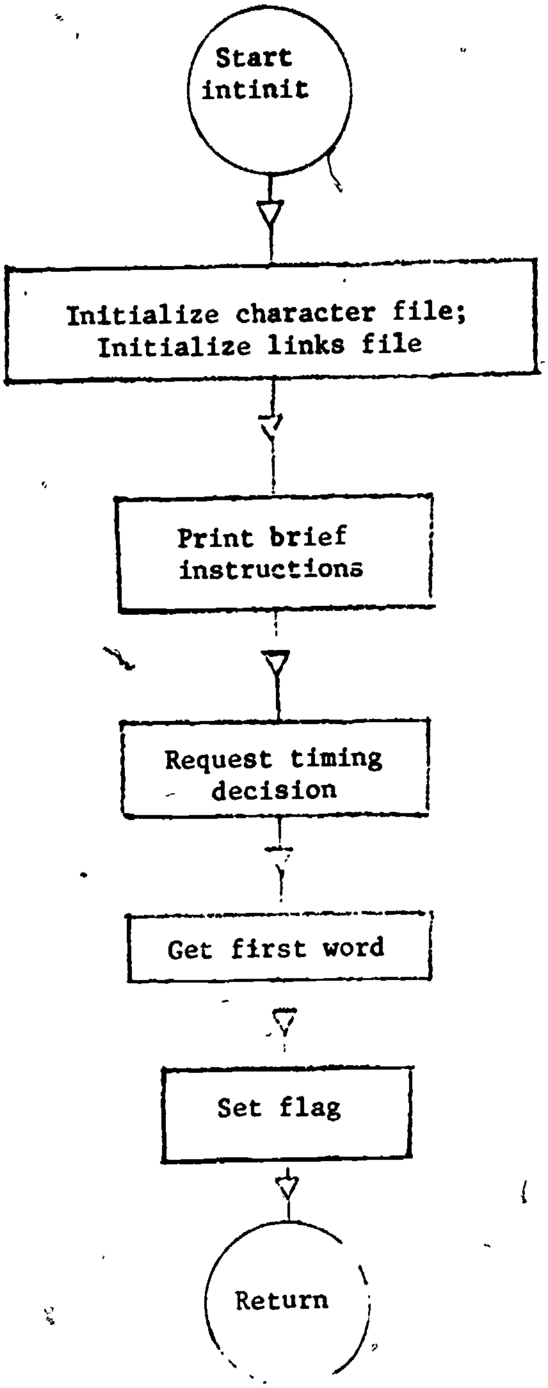


Fig. 85

EQL:
TEST FOR STRING EQUALITY

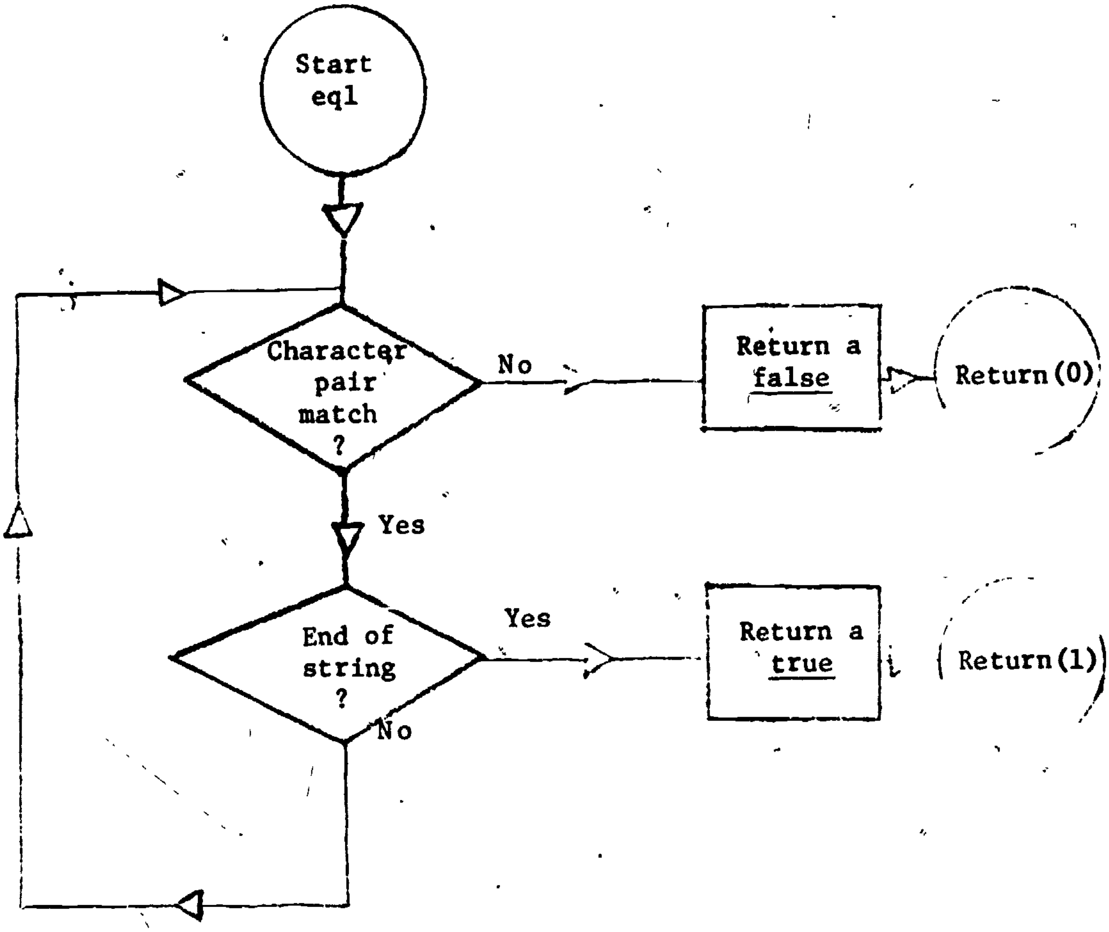


Fig. 86

GETSTR:
GETS A STRING WITH (POSSIBLY) EMBEDDED BLANKS

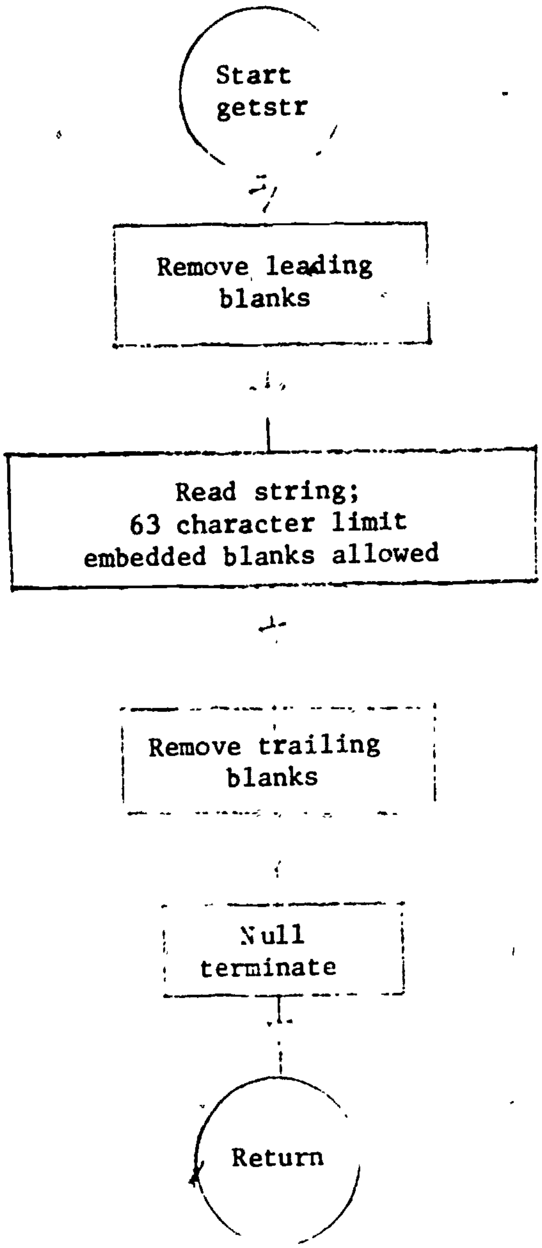


Fig. 87

X. REFERENCES

1. C Reference Manual. Western Electric Company (Bell Telephone Laboratories).

2. Thompson, K., and D. M. Ritchie. Unix Programmer's Manual. Fourth Edition. Bell Telephone Laboratories, Inc., 1973.

3. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, December, 1973. Progress Report, NSF Grant No. GN-36114.

4. Zunde, Pranas. Scientific and Technical Information Transfer for Education (STITE). Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, June, 1974. Progress Report, NSF Grant No. GN-36114.

APPENDIX

A. LETTER TO USERS OF THE GEORGIA INFORMATION DISSEMINATION CENTER (GIDC)

eorgia
stitute
of
echnology SCHOOL OF INFORMATION AND COMPUTER SCIENCE / (404) 894-3152 / ATLANTA GEORGIA 30332

November 26, 1974

Dear Colleague:

Most scientific and technical information resources that are available through computerized services (such as the service provided by the Georgia Information Dissemination Center of the University of Georgia) are utilized relatively little for educational purposes. Under a National Science Foundation grant to the Georgia Institute of Technology, we are currently studying the reasons why these information resources are so little used for educational purposes, and trying to determine ways and means of increasing their utilization for the improvement of instruction. In this study, we need to draw on the experience and expertise of people such as you, who have actually worked with systems of this type. We hope very much that you will be so kind as to help us.

Specifically, we are asking you and other users of the Georgia Information Dissemination Center to complete the attached questionnaire at your earliest convenience and return it to us in the enclosed pre-addressed envelope. Please feel free to supplement the questions on the form by adding your own comments and observations about your practices and experiences. Your assistance is essential to the success of this inquiry, from which the whole community of educators will eventually benefit.

Looking forward ~~to~~ your response, we thank you in advance for your kind cooperation.

Pranas Zunde
Professor
Information & Computer Science

PZ:tss

B. QUESTIONNAIRE TO USERS OF THE GEORGIA INFORMATION DISSEMINATION CENTER (GIDC)

1. Please indicate your position by checking the appropriate line.

_____ Undergraduate student
_____ Graduate student
_____ Instructor or lecturer
_____ Assistant Professor
_____ Associate Professor
_____ Professor
_____ Other academic (Please specify)
_____ Other non-academic (Please specify)

2. Have you used the information that you have received from the Georgia Information Dissemination Center (GIDC) for instructional purposes of any kind?

_____ Yes (Please go to question 3.)
_____ No (Please go to question 7.)

3. For what specific instructional or educational task(s) have you been able to utilize the computer-based bibliographic retrieval services of GIDC services?

_____ Development of a new course
_____ Updating of an existing course
_____ Preparation of illustrative examples
_____ Selection of case studies
_____ Compilation of bibliographies or reading lists
_____ Collection of data
_____ Current awareness in subject area of a course
_____ Preparation of quizzes, tests, and other exercises
_____ Other (Please specify)

4. In what subject area did you utilize this information? (For example: biology, chemistry, mathematics, physics, etc.)

____

5. For what types of courses have you used this information? (Please check as many as applicable.)

____ Lecture
____ Seminar
____ Special project
____ Laboratory
____ Other (Please specify) ____________________

6. Were titles and abstracts which were provided to you by the service sufficient for your purposes, or was it necessary for you to obtain full-text documents?

____ Titles and abstracts <u>were</u> sufficient
____ Titles and abstracts <u>were not</u> sufficient

7. What improvements to the system would enhance utilization for instructional purposes?

____ Easier access to the system
____ Shorter waiting times for information delivery
____ More descriptive abstracts
____ Browsing capability
____ Interactive system for query or profile formulation

8. What other suggestions would you have for the improvement of the system?

9. Eventually we might want to contact you for further comment and/or clarification. If that is agreeable to you, please provide the following information:

Name: ____________________

Telephone: ____________________