HOKKAIDO UNIVERSITY

Title	Construction of an Algorithm Information System ALGO
Author(s)	Amano, Kaname; Maeda, Takashi
Citation	北海道大學工學部研究報告, 117, 67-77
Issue Date	1984-01-31
Doc URL	http://hdl.handle.net/2115/41821
Type	bulletin (article)
File Information	117_67-78.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第117号（昭和59年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 117 (1984)

Construction of an Algorithm Information System ALGO

Kaname AMANO* and Takashi MAEDA**
(Received September 30, 1983)

Abstract

We have constructed an algorithm information system ALGO on an information retrieval system ORION, which is intended to collect, organize and store information concerning quality computational algorithms to be used over a wide range of scientific fields and to provide the information corresponding to requests of general users with various motivations. The algorithm information is represented in a conceptual framework consisting of the three basic sets of attributes, i.e., bibliographic, functional and operational ones, selected to describe the essential features of the computational algorithms. We have found that an information retrieval method is effective not only for dissemination of algorithm information but also for its organization and maintenance. However, two directions can be suggested for constructing a more integrated algorithm information system, i.e., one to the knowledge based system type and the other to the method base system type.

1. Introduction

Design or selection of computational algorithms is one of the most important steps in the process of problem solving by computers in scientific fields, and new algorithms often originate from modifications or revisions of existing ones. On the other hand, computational algorithms developed and used in scientific researches have been increasing in number with the rapid advance of computer technology and computational techniques, and they have been distributed through various media. It is not always easy for a general user of algorithms to find most suitable ones for his specific problem solving. Therefore, collecting, organizing, storing and providing the information about quality computational algorithms for vital use of scientific researchers in various fields are urgently needed.

We have constructed an algorithm information system ALGO (an ALGOrithm information system) on the information retrieval system ORION (Online Retriever of

* Hokkaido University Computing Center, Sapporo 060, Japan
** Department of Engineering Science, Faculty of Engineering, Hokkaido Uninversity, Sapporo 060, Japan

InformatiON produced by Hitachi Ltd.) [1,2] . By the algorithm information, we call collectively the information about algorithms of various computational fields (numerical analysis, data processing, non-numerical computations, ...) presented in various forms (executable modules, source programs, natural language texts with mathematical notations, ...) and the information about their analysis, performance evaluation, modification, revision, etc. The ALGO system takes up mainly the information about computational algorithms published in the form of source programs in scientific journals, some of which have been found to be the original information sources for the well known systems and packages of selected algorithms distributed by software suppliers.

The algorithm information with a complex information structure is represented in a conceptual framework consisting of the three basic sets of attributes, i.e., bibliographic, functional and operational ones, which are selected to describe the essential features of the existing computational algorithms. Each algorithm information represented in this framework is dealt with as an individual object by using the conventional techniques in the information retrieval system. In the ALGO system, a simplified version of the framework has been used as a first step of the progress. Some programs are stored on disks and available in the retrieval process.

This paper is concerned mainly with the construction of the ALGO system. See the companion paper [3] for more detailed contents and usage of the ALGO database.

2. Algorithm Information

The first problem for constructing an information system is how to represent and describe the essential contents of the objective information to be treated. The algorithm information is a kind of procedural information with a complex structure reflecting the following characteristics of the computational algorithms: difficulty of the definition of their elements, diversity of the existing forms, changeability by the subsequent revision or modification, flexibility for applications to various types of problems, etc. On the other hand, user's needs are generally more selective than that in the usual document retrieval. They want a few algorithms, typically only one algorithm best suited for their specific problem solving. Thus, it will be practical to start from dealing with published algorithms independently for avoiding difficulty caused by the reorganization of various algorithms with a single formal language as a whole.

We have employed a conceptual framework to represent the information on individual algorithms, which consists of the following three types of attribute-value sets intended to describe their essential features:

(1) bibliographic attributes containing the information items adopted in the usual document retrieval such as title, authors, journal, volume, number, page, year, etc.,

(2) functional attributes representing the essential features of their functional contents or properties, which consists of the basic three items of the applicable prob-

lem domain (object and operation), method (name and techniques used) and performance (complexity, storage requirement, accuracy and other conditions), and

(3) operational attributes describing the information items needed for reference to or execution of the existing programs such as programming language, usage, running environment, histoy, etc.

Their values are mostly represented by technical terms in computional mathematics or computer sciences. This framework, which gives a functional indexing method analogous to that of more content-oriented document retrieval [4] , is generally applicable to the computational algorithms of various scientific fields and existing forms. And the essential features of the algorithm information can be described for its utilization in the information retrieval.

One of the most important problems has been how to extract the features which, in some systematic way, classify problems and determine the applicability of the algorithms to them [5-7] . We could define several subschemes, each for a specific computational field, if the fields are well defined. For example, in the field of simultaneous linear equations, the subscheme should have the attributes such as problem size, symmetriness, bandedness, diagonal dominance, conditioning, positive definiteness, scaling, etc. However, at the present stage, the general scheme will be more practical for our purpose.

3. ALGO Database

Information sources of the ALGO database are ACM algorithms, Algorithms Supplement of The Computer Journal which we call CJ algorithms, Index by Subject to Algorithms appearing in the Communications of the ACM collected from ten to twenty scientific journals and publications, and research papers of the ACM Transactions on Mathematical Software (Table 1). The Index by Subject to Algorithms includes ACM and CJ algorithms but, for the early years of 1960-1980, only the brief information of an abbreviated form is recorded. These algorithms cover a wide range

Table 1 Information sources of the ALGO database

	ACM Algorithms	(1960-1982)	1045	(591 algorithms)
	Comm. ACM			
*	ACM Trans. Math. Softw.			
	ACM Trans. Program. Lang. Syst.			
	CJ Algorithms	(1966-1982)	159	(114 algorithms)
	Computer J.			
	Index by Subject	(1960-1980)	1586	
01		(1981-1982)	57	
	Comm. ACM			
	Papers	(1975-1982)	200	(200 papers)
	ACM Trans. Math. Softw.			

* unit : records (≈algorithms)

Table 2 Field definition of the record of ALGO database

no	name	type		prefix		output label	contents
1	DN	S KEY				D DOC NO	document number
2	AN	S	INDEX	AN:		D ALG NO	algorithm number
3	TI	S	INDEX	TI:	KE:	D TITLE	title
4	AU	S	INDEX	AU:		D AUTHOR	authors (affiliation)
5	JO	S	INDEX	JO:		D JOURNAL	journal
6	VO	S				D VOLUME	volume
7	NU	S				D NUMBER	number
8	PA	S				D PAGE	page
9	YE	S	INDEX	YE:		D YEAR	year
10	RE	S				REFERENCE	references
11	SH	S	INDEX	SH:	KE:	D SHCODE	SHARE classification cedes
12	CA	S	INDEX	CA:	KE:	D CATEGORY	ACM CR categories and subject descriptors
13	GT	S	INDEX		KE:	D G TERM	ACM CR general terms
14	KE	S	INDEX		KE:	D KEYWORD	additional keywords and phrases
15	AB	S				ABSTRACT	abstract
16	PL	S	INDEX	PL:	KE:	D P LANGUAGE	programming languages
17	SU	S	INDEX	SU:		D BY SUBJECT	index by subject to algorithms (1960-1980)

of computational fields and have been used widely.

There are various types of documents for the algorithms: algorithm (proposal of an algorithm in a programming language with brief explanation about its function, performance, usage, ...), certification (a report on a previously published algorithm about its correctness, extensive testing, ...), remarks (a report on a previously published algorithm usually concerned with correction or modification), and the others including notes, translations, surveys, etc.

The ALGO database consists of records, each corresponding to a document, and the record consists of fields (Table 2) roughly corresponding to the attributes described before. The fields DN-RE, SH-AB and PL are for the bibliographic, functional and operational attributes respectively, though the description for the functional attributes are not separated into the problem domain, method and performance. The data of Index by Subject to Algorithms (1960-1980) has only the two fields of DN and SU; the latter is not shared by other data and contains an abbreviated form of information corresponding to AN, TI, JO, VO, PA, YE, SH and PL. The field types S and KEY mean the character string and the unique accession number, respectively. Index terms are extracted from the fields denoted by INDEX. Note that the fields with two prefixes are doubly indexed and a virtual field with the prefix "KE:" contains the whole functional information of the algorithm. The symbol D in the output label denotes the default fields for display.

4. ALGO System Organization

4. 1 Information Retrieval System ORION

Fig. 1 shows the general organization of the ALGO system.

ORION [8,9] is an information retrieval system with inverted files and can deal with the data types of both characters and numeric values. Three groups of functions are prepared in the system (Fig. 2) : (1) data definition by the DDL (Data Definition Language) and database construction and updating by input data preprocessing, automatic extraction of the index terms, etc., for the database administrators, (2) system management for the database system managers, and (3) retrieval by index and sequential searches with various search conditions and auxiliary retrieval assistances such as using profiles (a stored set of commands), requesting information about the

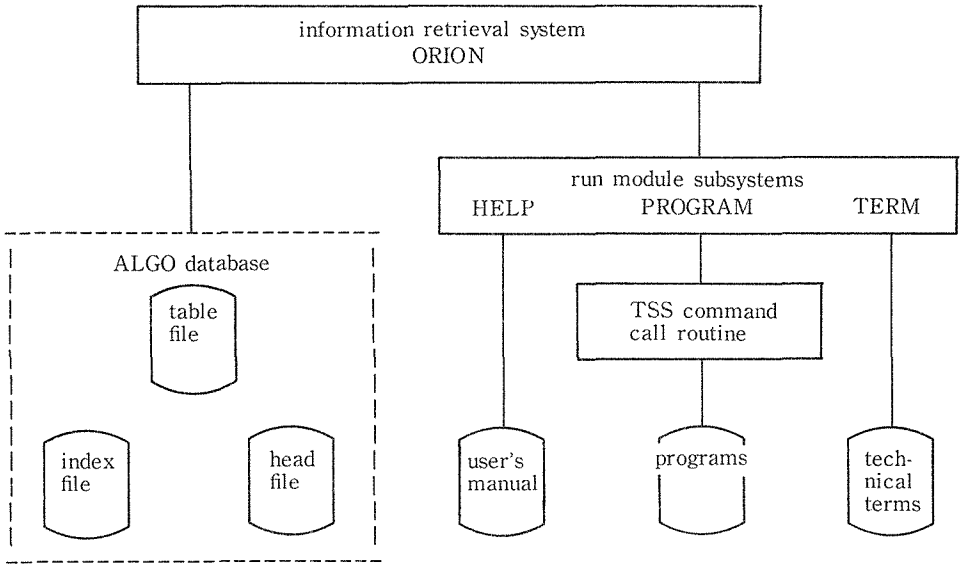

Fig.1 General organization of the ALGO system

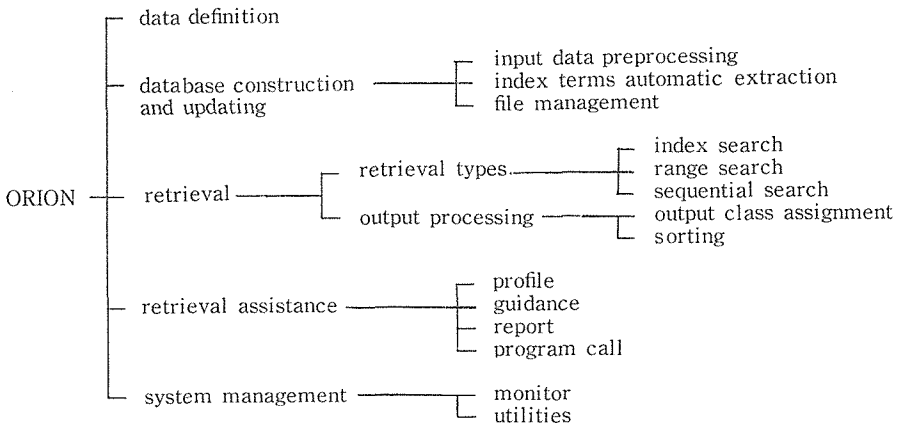

Fig.2 Function tree of the ORIONsystem

system and its usage, report making, calling programs called run modules, etc., for the general users.

The ALGO database consists of the following three files: (1) table file which stores the information about the definition of files, records, fields, security, etc., (2) head file which stores the ALGO database records themselves, and (3) index file which stores index terms and pointers to the corresponding records [Appendix A1].

The work procedures of database construction and updating are generally very complicated [Appendix A2]. However, it has been simplified to a large extent by a set of command procedures prepared [10].

The major commands for searching data are FIND for the high speed index search and SCAN for the sequential search with various search conditions for the contents of the specified fields of the specified set of records. Index terms, together with the prefixes of the corresponding fields, are available to describe the search conditions of the index search. The contents of the set of the hit-records resulting from these searches is displayed by the command DISPLAY.

4. 2 Run Module Subsystems

The ALGO system has been greatly enhanced by the funciton of the run module call which makes one possible to combine the ORION system and the external programs. We have implemented the following subsystems in FORTRAN and PL/1.

(1) HELP subsystem has the following functions based upon the user's manual stored on disks: the online reference to information such as a table of topics available, information sources and contents of the ALGO database, definitions of the fields of the records, classification codes, usage of the main commands, and the offline print of the manual.

(2) PROGRAM subsystem is for listing and copying of programs stored on disks, i.e., FORTRAN, PL/1 and some ALGOL programs of the ACM algorithms. This subsystem has an interface to the TSS command execution routine [11], then almost all processings by the HITAC VOS 3 TSS commands including compiling, linking and running are available, in principle. But, users must, in advance, make all preparations of data, driver routines, etc.

(3) TERM subsystem is based upon the online dictionary of about 11 thousand

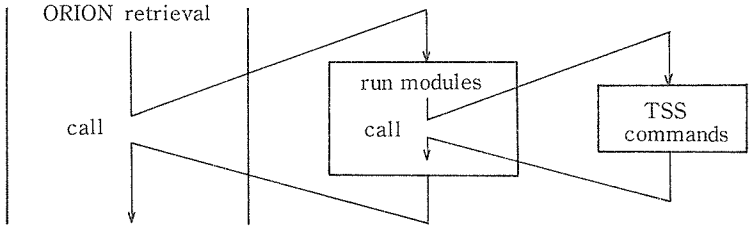

Fig. 3 Control flow of the run module call

technical terms in mathematics and computer science collected from several mathematical dictionaries. The following types of information are available : correct spellings of the technical terms and their related terms, their information sources, etc.

Fig. 3 shows the control flow of the run module call. The names of the subsystems HELP, PROGRAM and TERM are the command names available in the retrieval process.

5. Brief Example of the ALGO Retrieval

Fig. 4 shows an example of the ALGO retrieval, which proceeds typically with the following pairs of input

```
WHICH TOPIC ?
SHCODE
  :
F4  SIMULTANEOUS LINEAR EQUATIONS
E1  INTERPOLATION
  :

   1/ FIND SH:F4
*      57    1/  SH:F4
   2/ FIND LINEAR AND EQUATION*
*     132    2/  LINEAR
*     152    3/  EQUATION (2 TERMS COMBINED)
*      54    4/  LINEAR AND EQUATION*
   5/ FIND 1 AND 4
*      25    5/  1 AND 4
   6/ FIND 4 AND SYMMETR*
*      31    6/  SYMMETR (2 TERMS COMBINED)
*       3    7/  4 AND SYMMETR*
   8/ SCAN  7 TI POSITIVE ADJ DEFINITE*
*       1    8/  SCAN  7  TI POSITIVE ADJ DEFINITE*
   9/DISPLAY

DOC NO       934
ALG NO       ACM ALGORITHM 512
TITLE        A NORMALIZED ALGORITHM FOR THE SOLUTION OF POSITIVE
             DEFINITE SYMMETRIC QUINDIAGONAL SYSTEMS OF LINEAR
             EQUATIONS
AUTHOR       BENSON A. AND EVANS D. J. (LOUGHBOROUGH UNIVERSITY OF
             TECHNOLOGY, ENGLAND)
JOURNAL      ACM TRANSACTIONS ON MATHEMATICAL SOFTWARE
VOLUME       3
NUMBER       1
PAGE         96-103
YEAR         1977
SHCODE       F4
CATEGORY     5.14, 5.17
KEYWORD      LINEAR EQUATIONS, NORMALIZED SOLUTIONS, PERIODIC
             QUINDIAGONAL, SYMMETRIC POSITIVE DEFINITE
P LANGUAGE   FORTRAN

   9/
```

Fig. 4 Brief example of the ALGO retrieval

i/ FIND OPERAND
i/ SCAN j FIELD OPERAND

and output

* k i/ [j FIELD] OPERAND.

Underlines show the input part. The symbol "i/" in the input line is the prompting from the system for a command; "i" becomes the set number of the "k" numbers of hit-records resulting from the corresponding search request, and increases by 1. OPERAND stands for the search condition, accompanied with the assignment of the set number "j" and the field name in the case of sequential search. If no prefix is given in the operand of the index search, "KE:" is adopted by the system. The symbol "*" in the operand terms such as "inver*", "symmetr*" and "definite*" is used for the part to be disregarded in the matching of the indicated character string. Set operations are available by the index search. The set of the number 8 containing only one record of the ACM algorithm 512, which is displayed by the "9/ DISPLAY", results after all from the following search condition: One of the fields with the prefix "KE:" has the terms of "linear", "equation*" and "symmetr*", and the title has the phrase of "positive definite*".

6. Concluding Remarks

We have constructed an algorithm information system based upon the information representation scheme introduced to describe the essential features of the algorithm information. The system has been examined for some typical requests and the information retrieval method has been shown to be used effectively not only in dissemination of the algorithm information but also in its organization and maintenance. However, the two following directions can be suggested as possible ways to improve the present ALGO system to a more integrated and better algorithm information system.

(1) First is the way to a knowledge based algorithm information system. A simple set of discrete classification codes and keywords with role indicators are not always sufficient to describe functional contents of the computational algorithms, especially for selective and critical retrieval for a specific problem solving by a general user. Mathematical facts and expert heuristic knowledge, which we call the algorithmic knowledge [2], are also important and should be collected, organized, stored and provided for vital use of scientific researchers. Some knowledge representation languages and techniques in artificial intelligence researches will be applied to specific description of the functional attributes and effective utilization of the algorithmic knowledge.

(2) Second is the way to an algorithm information system of the MBS (Method Base System) [12] type, which is a highly integrated system consisting of databases and

algorithms permitting immediate application of a selected algorithm to a specified data. It should be noted that the performance of the computational algorithms must often be evaluated by experimental comparisons among similar ones applied to a specific problem, then the function of the immediate application of a selected algorithm to a specified date is also important for the selection of the algorithm itself. The DBMS (DataBase Management System) will be needed from a view point of its ability of data management rather than that of information retrieval.

Acknowledgement

The authors wish to thank Mrs. A. Mochida and Mrs. Y. Nakayama (Hokkaido University Computing Center) for their kind help in construsting the database and the system, and to Mr. M. Chiba (Hokkaido University Computing Center) for his useful discussion at an early stage of the progress of this work. Thanks are also due to Prof. M. Kitamura for his careful and patient reading of the manuscript.

References

[1] Amano, K., Chiba, M., Mochida, A. and Maeda, T.: Proc. JPL/ACM-SIGNUM Conf. on the Computing Environment for Mathematical Software, Pasadena, 8−9, 1981. and also ACM SIGNUM Newsletter, Special Issue Oct. 1981, 8−9.

[2] Amano, K., Chiba, M., Mochida, A. and Maeda, T.: submitted to Information Systems.

[3] Amano, K. and Maeda, T.: Bull. Faculty Engrg., Hokkaido Univ. in this issue.

[4] Maeda, T: Inform. Proc. Management, **17** (6), 329−339, 1981.

[5] Rice, J. R.: Mathematical Software (ed. Rice, J. R.), 13−25, and 27−41, Academic Press, 1971.

[6] Rice, J. R.: Advances in Computers 15 (ed. Rubinoff, M. et al.), 65−118, Academic Press, 1976.

[7] Rice, J. R.: Performance Evaluation of Numerical Software (ed. Fosdick, L. D.), 301−307, North-Holland, 1979.

[8] HITAC VOS3 manual (8090-6-502), Hitachi Ltd. (in Japanese).

[9] HITAC VOS3 manual (8090-6-503), Hitachi Ltd. (in Japanese).

[10] Mochida, A., Sagara, T., Sugiura, T., Kaida, T. and Amano, K.: in preparation as a user's manual for database construction, Hokkaido Univ. Computing Center. (in Japanese).

[11] Adachi, J.: Center News, Tokyo Univ. Computing Center, **14** (11), 96−100, 1980. (in Japanese).

[12] Bever, M., Gaube, W., Lockemann, P. C. and Mayr, H. C.: Database Management: Theory and Applications (ed. Holsapple, C. W., et al.), 107−130, D. Reidel Pub. Co., 1983.

Appendices

A1. ORION File Organization

Fig. A1 shows the general organization of the ORION files. The database consists

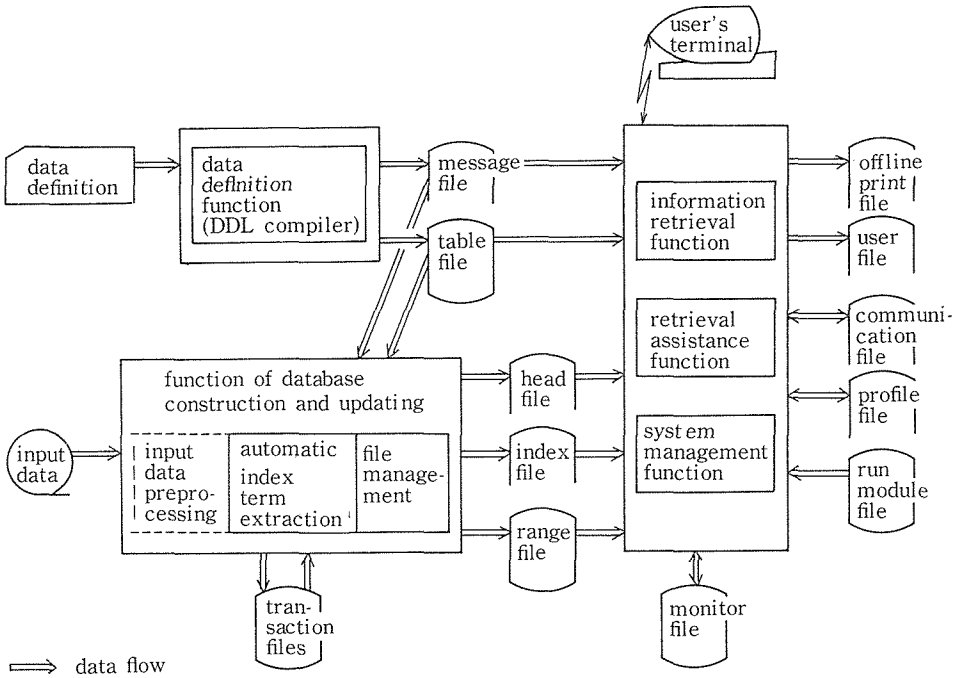

Fig. A1 File organization of the ORION system

of a table file (definition of files, records, fields, security, ...), a head file (database records themselves), an index file (index terms and the pointers to the records) and a range file (range terms and the corresponding numerical values). The message file contains the ORION messages. The communication file (work file for the retrieval results), the offline print file (information for the offline LP print) and the monitor file (information monitored in the ORION session) are allocated in each ORION session. The profile file (a set of retrieval commands) and the user file (the retrieval results) are for auxiliary retrieval assistance. The input data file (original data) and transaction files, i.e., head file transaction (data in the ORION format to be stored in the database), index file transaction (indx terms and the information of their corresponding records), range value transaction (numerical data) and range file transaction (values for each range term), are used for database construction and updating. The run module file stores the ORION's and user's run modules.

A2. ORION database construction and updating

Fig. A2 shows the work flow of the ORION database construction and updating. The parts enclosed by dashed lines are used for updating.

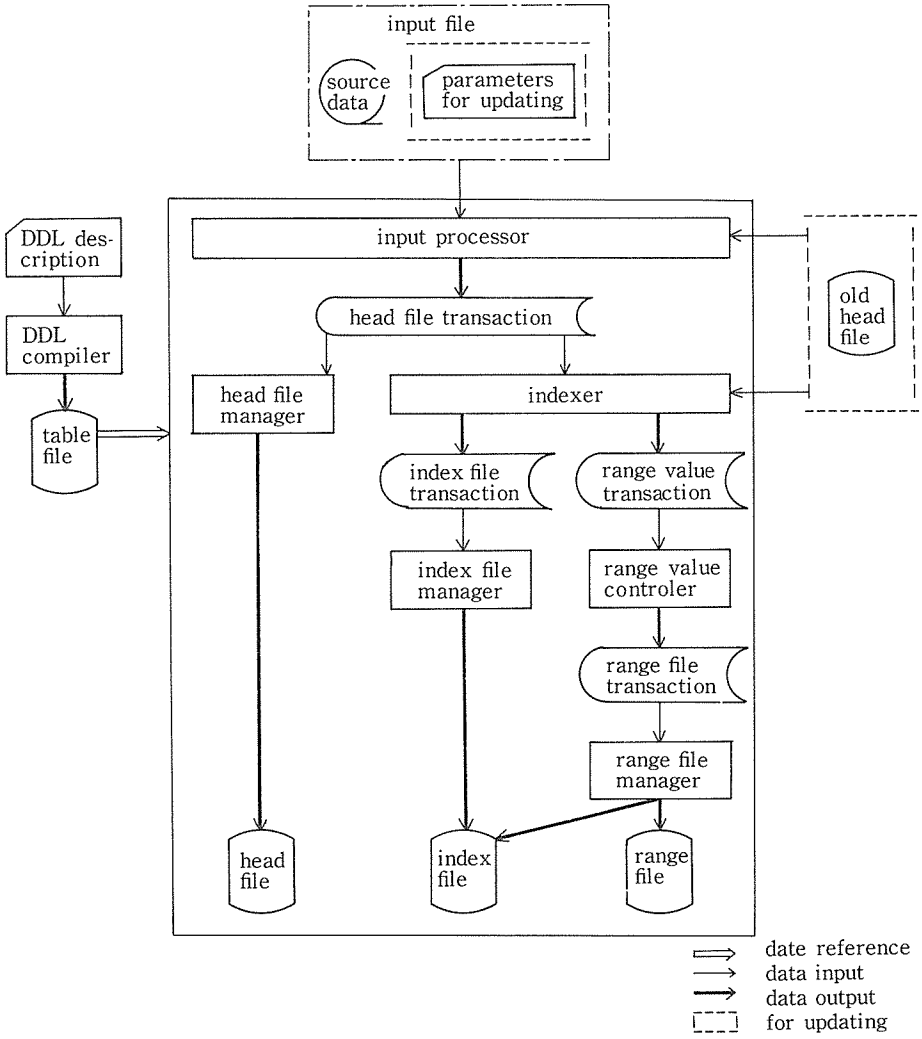

Fig. A2 Work procedure of the ORION database construction and updating