Focus

VALUE-ADDED GLOBAL DATABASE SEARCH THROUGH WIDE AREA NETWORK*

Anjali Baokar, PR Sundareshwar and MR Joshi
Research & Development Establishment (Engineers), Pune 411 015.

Abstract

Online database searches have become common and a number of American and European databases, notably those related to science and engineering have become very popular due to their specialisation in specific fields. Although telecommunication charges are relatively low, specific database charges are substantial, depending upon the area of specialisation. This paper discusses the access and search of global databases, value addition through organisation of the received data and sharing of the organised data in the local area network. The original organised database when accessed and downloaded becomes an ASCII file. If the selected database is to be reconstructed at the receiving end, it is a laborious task of manual data entry which could be very slow and error-prone. A program developed for automated conversion of the ASCII file into a standard database format is described.

1. INTRODUCTION

A database may be defined as a collection of interrelated data stored together without redundancy to find multiple applications. It is, in fact, an array of data organised in a logical manner with minimum functional and maximum tolerable redundancy, indexed in a way most optimal for the kind of application the user is interested in. This data can exist on a variety of physical media, such as card files, magnetic tapes, etc.[1]

Database environment can be classified into operational environment and decision support environment. The primary aim of the operational environment is towards online performance and data integrity, whereas the decision support environment is aimed at user flexibility in viewing the data. The decision support environment operates on downloaded data that is

* *Reprinted with permission from Defence Science Journal 1993, 43(2).*

summarised or data that has been screened, including editing, merging, recalculating, etc.[2]

Global databases come under decision support environment. They can be defined as a collection of information which is accessed all around the globe. These databases cover practically all fields of human activity. The databases are built through contributions from several universities and organisations. They form the most important part of today's information technology and offer almost all the needed information instantaneously. They support a large number of queries made by today's R&D organisations and industries.

These global databases are accessed by the users all around the globe through public data networking using satellites. This accessed information can either be received through courier services or can be downloaded to a user-end computer. These global databases are viewed as valuable resources. Access to these databases is charged, depending on the value of the database, satellite time used, etc.

Databases searches usually serve a useful purpose immediately, but paper copies tend to get lost. If the same information is downloaded into computer format, it is preserved for a longer duration, but cannot be retrieved easily, since it is not in the form of a structured database. A study has, therefore, been carried out to detect the received information format and suitable identifiers located so as to facilitate structuring of the information with minimum manual effort.

A cost-effective way of searching these databases and sharing the facility among different Defence establishments is presented in this paper. We propose that individual subscribers to these databases, after search, shall maintain the searched data in a structured and sharable form at their work place. Later, these users shall send their searched data to a commonly identified nodal agency in the premises. At the time of any new search, the users shall first search in the common database maintained at the nodal agency. If the query is not answered, then they are advised to search the global database. As most of the defence establishments are subscribing to DIALOG DATABASE SERVICE, we present a case study on DIALOG.[3]

2. DIALOG DATABASE

2.1 What is DIALOG

DIALOG is an 'electronic library' used to locate information on most of the subjects from a variety of sources. Using DIALOG, we can quickly browse through 'volumes' of information.

The DIALOG system manages a wide variety of information, including references and abstracts relating to published literature, statistical tables, directories, business and financial data (such as financial reports), and for a rapidly growing number of publications, the complete text of articles.

This information has been collected, organised and produced in machine-readable form (on magnetic tape) by many publishers, who, in turn, are contracted by DIALOG to store their collections on its mainframe computers.

DIALOG customers can search for and view the information contained in over 300 databases via telecommunication networks. A single database contains anywhere from 3,000 to 70,00,000 records, which can be browsed online, downloaded on to local disk or printed offline.

2.2 How to Access DIALOG

Hardware

IBM or compatible PC/AT using a modem, compatible with CCITT Standard with 300, 1200, 2400 or 4800 baud rate, which can be internal or external to the computer, and telephone lines/packet-switched data networks.

Software

Communication software like DIALOGLINK, PROCOMM, etc.

To access DIALOG the user first logs into the national networks, which, in turn, hooks to international networks, via GPSS, Bombay.

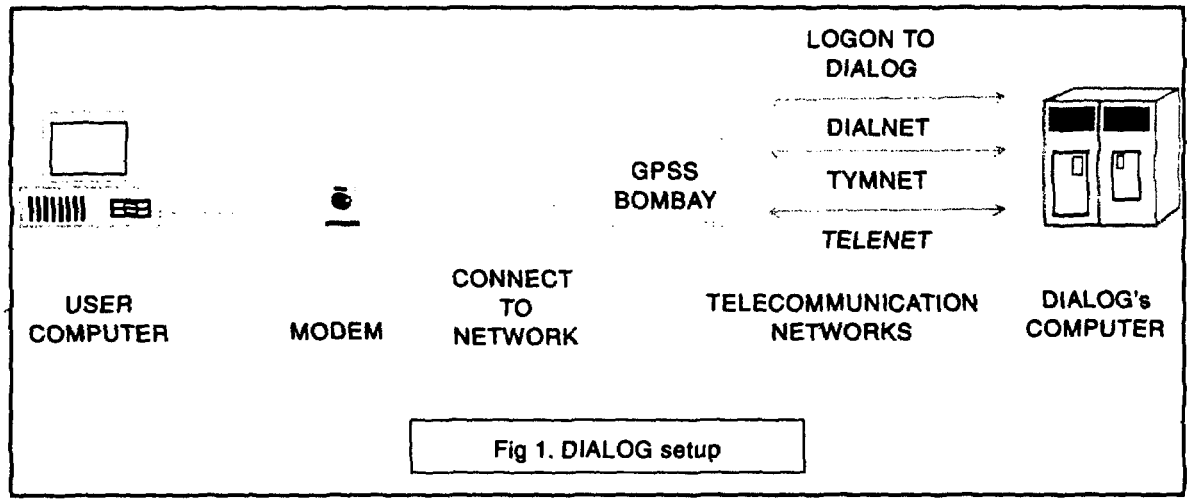

Fig 1. DIALOG setup

DIALOG is accessible through the following US telecommunications networks:

- DIALNET and INWATS
- TYMNET
- TELNET

The set-up required to access DIALOG, is shown in Fig.1.

2.3 Billing Methodology

DIALOG has an annual renewal fee per password. There is no monthly minimum. A DIALOG subscriber pays only for the time he actually uses this service.

Factors affecting monthly invoice include

(a) Connect time:
Charges for the duration of time spent online. Charges vary with different databases.

(b)Telecommunication charges:
Charges for the telecommunication network used.

(c) Per-item type charges:
Charges for viewing records accessed during the search.

General Format for DIALOG Search

The following are the pre-defined formats for DIALOG search

Format 1 : DIALOG accession number

Format 2 : Full record except text

Format 3 : Bibliographic citation

Format 4 : Full record with tagged fields

Format 5 : Full record

Format 6 : Title and DIALOG accession number

Format 8 : Title and word count

3. RE-STRUCTURING OF SEARCHED DATA

The downloaded data at the user end is an unstructured ASCII file created depending on the search carried out. This ASCII file should to be re-organised and structured. Then this file is ready for sharing among different users as a database.

A method for structuring the downloaded file, is presented in the following sections.

3.1 Underlying Concept

The search is carried out by specifying the search number followed by search format and number of records. The records are separated by this key in the received text.

key : i/j/k

where i is the search number, j is the search format, and k is the record number.

unstructured ASCII file has to
ss of re-organising and structur-
this reorganisation is explained
... ... e following algorithm.

3.2 Algorithm

read input file

search for key in text file

if key is found

do while not of input file

repeat

copy the token to temporary file

read next token

until next token = key

end do

read keywords and subject for classification

build appropriate database based on subjects from keywords and temporary files.

remove all temporary files.

The flowchart of this algorithm is presented in Fig. 2.

3.3 Implementation at R&DE (E)

Research & Development Establishment (Engrs) (R&DE(E) is a subscriber to DIALOG database. In addition, R&DE(E) has a local area network spread across its campus. The data searched is downloaded and restructured, as described above. This database is added as a new resource to the library information system on the local area network and is now being shared among the users of the network. Within a short time, the number of records available in R&DE(E) database has crossed 500.

4. PROPOSED SYSTEM

4.1 Value-Addition through Sharing using WAN

When we have different subscribers of DIALOG situated at different places in India carrying out searches in various fields, it is appropriate to mutually share the databases.

These individual subscribers will first structure their searched information, reorganise it to form a shared database, and send the same to a common nodal agency. This nodal agency shall be responsible for maintaining a common database. Later on, a user will first access the common nodal agency to find the required information. If he does not get the needed information, he will go for DIALOG access.

Thus, value is added to this unstructured received data by developing it into a database and sharing the same among many users. The user will be paying only the communication charges in case he gets the needed information from the nodal agency. If the search fails, he has to pay only nominal overhead charges. The main advantage is that this common database will grow with time and will be able to support more queries.

The value-addition is in the most important area, i.e., cost. Cost-effectiveness becomes obvious from the following comparison.

When Searched DIALOG using international network

Total charge = DC+TC

When Searched at nodal agency using national network

Total charge = TC

where DC is the DIALOG charges (connect time charges + per item type charges), and TC is the telecommunication charges (call duration charges + data transfer charges).

Further, the telecommunication charges for a national call are one fourth of those of an international call.

Example

Consider a typical sample DIALOG search carried out on a certain topic using NTIS file (File number 6) which results in a collection of 100 records on that topic. The expanses incurred are:

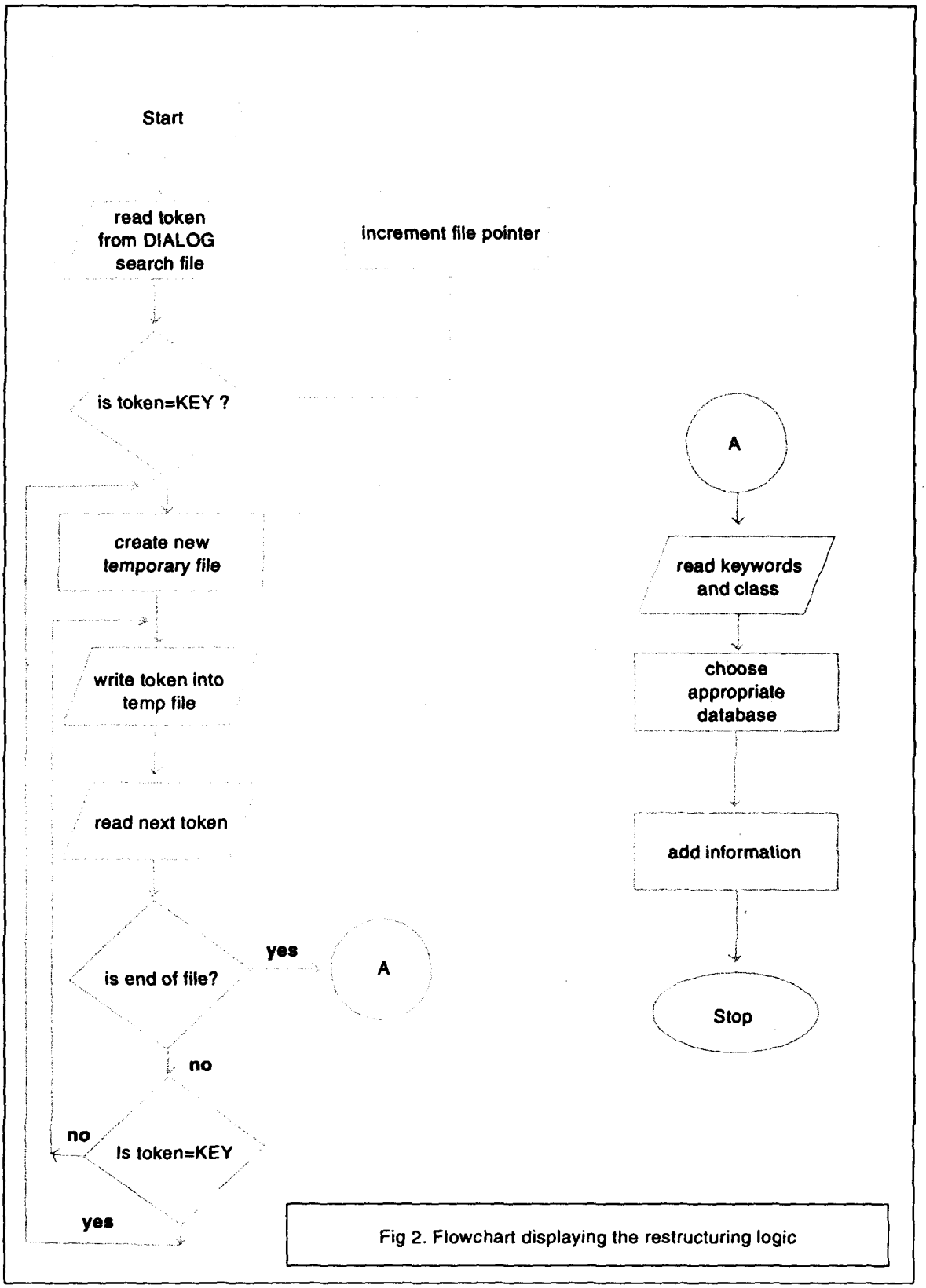

Fig 2. Flowchart displaying the restructuring logic

Telecommunication Charges	
Initialising & dialling	10 min
Searching	10 min
Type/downloading	50 min
(1/2 min required for 1 record)	
Total communication time	70 min
Communication cost	Rs 280
(Rs 4 per min)	
Cost of data transfer	Rs 606
Total telecommunication charges	Rs 886
DIALOG charges	
Searching	10 min
Downloading	50 min
Total connect time	60 min
Connect time cost	$ 84
Print charges/100 records	$ 65
($ 0.65 per record)	
Total DIALOG charges	$ 149

The total cost incurred on DIALOG search = Rs 886.00+ US $ 149.00= Rs 5356.00 (@ 1 US = Rs 30).

The total cost estimated for a successful search among the users connected through national network = Rs 221.50

4.2 Sharing Using INET (Packet-Switched Data Network)

Various Defence Establishments including R&D units, services, production units, etc are located all over the country. They all need information about Defence technology, but in different contexts. Some of these Defence units are subscribers to global databases. The information searched by these units may at times be redundant, resulting in duplication of search and hence extra cost.

We propose that a nodal agency be identified to maintain the database, constructed through contributions from different subscribers. Figure 3 shows how the DRDO database will grow richer through multiple users. The subscribers will first search through this database maintained by the nodal agency and then search in DIALOG if the search at the nodal

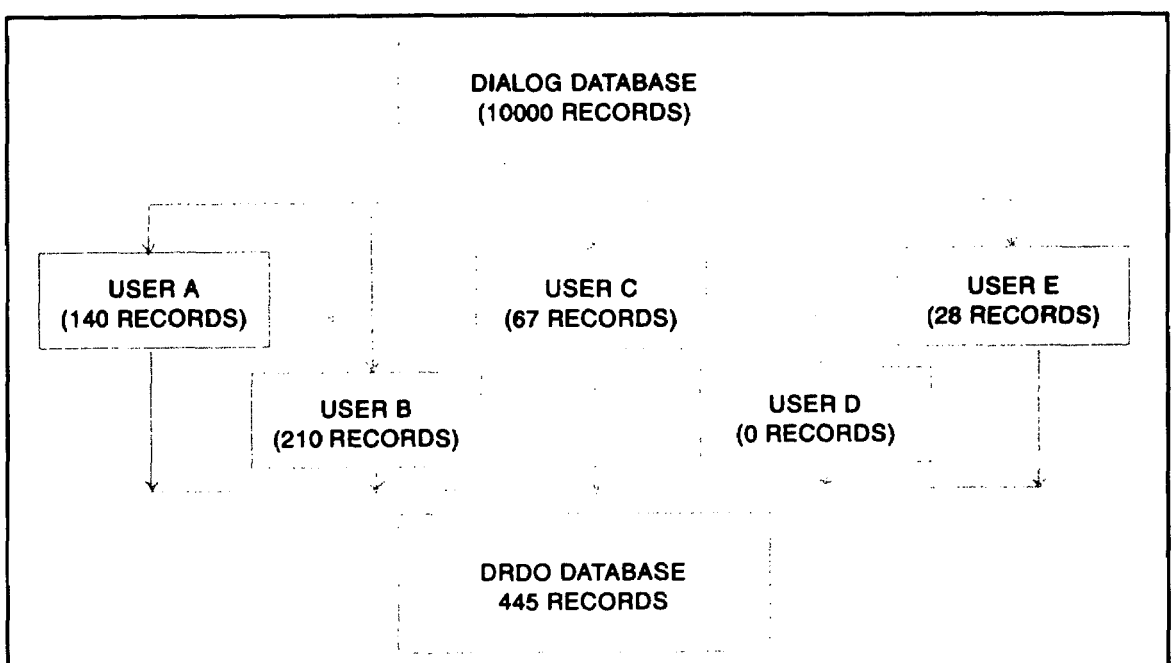

Fig 3. Growth of DRDO database through multiple users

agency fails. Also, the users, when searching through DIALOG, will add the searched information to the database maintained by the nodal agency. The individual users will be accessing this common database maintained at the nodal agency through I-Net, introduced by the Department of Telecommunications.

I-Net offers facilities for interlinking terminals and computers through packet switches located in important cities of the country. Subscribers to this network can also access international networks abroad through GPSS of VSNL at Bombay.

Although expensive, access to global databases is required to keep R&D personnel abreast of advances in various fields of their interest. A cost-effective way is to share the information maintained at the nodal agency, eliminating redundant searches of global databases.

4.3 Legal Aspects

The proposed information sharing is legal, since the information stays within the customer's organisation and is not resold.

REFERENCES

1. Henry, F. Korth & Silberschat, Abraham. Database system concepts. Mc-Graw Hill, Singapore, 1986. pp 422-44.
2. Inmon, W.H.& Thomas,Jr. J.Bird. The dynamics of database. Prentice-Hall, New Jersey, 1986. pp 17-20.
3. Searching DIALOG: the tutorial guide. DIALOG Information Services, March 1988.

"The art of life, indeed consists largely in the capacity to spend wisely and happily the moments in which we are most free—free from the demands and behests of others, free from the demands of relaxation and hygiene, free to plan our activities in and at our own time. To waste these moments is to waste something extraordinarily precious."

-SR Ranganathan