MODULARITY, MICROFORM & MICROCOMPUTERS

J. W. Cherry & E. V. Goodman
Exxon Corporation
New York U.S.

ABSTRACT

When after over fifteen years' experience in developing and operating Information Retrieval systems Exxon Corporation developed its fourth generation system, LIIMS, it built a high degree of modularity into the system. One of the major modules, the Index Management Module, can be regarded as consisting of up to seventeen sub-modules. The user can select the sub-module or sub-modules which best meet his specific needs and vary this selection from task to task without penalty or programming demands.

This modularity has made it possible for Exxon to develop a Microform Information Transfer Center and exploit the new microcomputers with minimal systems effort. Searches of COM-produced indexes in the MITC lead to the recovery of the full-text of appropriate documents. The use of microcomputers has speeded up the indexing process, eliminated key punch and greatly facilitated editing, thesaurus construction and on-line printout. Equally exciting, it has simplified access to external information banks.

MODULARITY, MICROFORM & MICROCOMPUTERS

Exxon's Logically Integrated Information Management System (LIIMS) owes a great deal to the experience Ed and I gained developing and operating three Information Retrieval Systems--SIS I, SIS II, and SIS III, which originated in Canada. When we were assigned the task of developing LIIMS, we were in the fortunate and possibly unique position, for people working in the information field, of being invited to develop a fourth generation system after over twelve year's experience of working with computer-orientated information systems.

All of this experience suggested two things. First, that our fourth generation system should be user-oriented; second, that it should be highly modular in construction.

Let me address myself to the first of these premises. SIS I in its day was a fairly advanced system. We could make it perform a number of useful tasks. However, in Imperial's Technical Information Services Department, we had to maintain a "resident specialist" who made out complicated run-sheets for every piece of output we requested from the system. During the "up-date and run" period of our routine operation, he was in constant communication and often, conflict with a "resident specialist" in our Systems and Computer Services Department. SIS II and SIS III, to the best of our knowledge, did little to remove this barrier to system exploitation.

In the Index Management Module of LIIMS, we have removed this barrier completely. The user of the system, electing to choose between two mini-program decks, can run a pilot program, or have the computer produce any index to his specific requirements, directly. We can train him to use these mini-decks in half a day, and then he is completely independent.

We developed this approach before we began to use our micro-computer. I should emphasize, in fairness to Ed who designed the front end programs that made it possible, that simplifying the user/system interface calls for some very sophisticated programming, and that the introduction of these programs was a pretty traumatic "happening" for the client-run coordinators whose role in life up to that time was to act as a human interface between the user and the system. At the same time, I should point out that his approach is now being universally adopted by our Maths, Systems, and Computer Department, and that it greatly facilitated our move to the micro-computer.

Now let me address myself to the second premise - a modern IR system must be modular in design.

If anything confirmed our early suspicions about modularity, it was working with SIS II. SIS II, as I hope most of you know, was a

co-operative effort of Imperial, the Canadian Geological Survey, the U. of Saskatchewan, and the U. of Calgary. The programs were written by the U. of Saskatchewan and have always been regarded as extremely effective programs. The trouble with SIS II as a system however, is that if you want to make any changes to its routines or its output, you almost have to take it to pieces, modify it, and then put it back together again. If we learned anything from SIS II, it was that you had to be able to modify without dismembering - if you wanted a viable system.

At the same time as we were beginning to realize this basic fact, our ideas about "Information Banks" which we had first discussed at a meeting at the U. of Manitoba in September 1967, were beginning to crystallize and force us into modularity.

We define an "Information Bank" as a discrete collection of information maintained, organized and accessed with the assistance of a computer, pragmatically separated from other collections of information maintained, organized and accessed by the system because of differences in discipline, vocabulary, usage or even client preference.

If you think about that for a moment, you will realize we are inferring that even input to an IR system should be modular.

With all this in mind, we built seventeen modules into the Index Management Module of LIIMS.

The Index Management Module itself may be regarded as one of five major modules of LIIMS. Figure 1 shows the IMM modules in summary form.

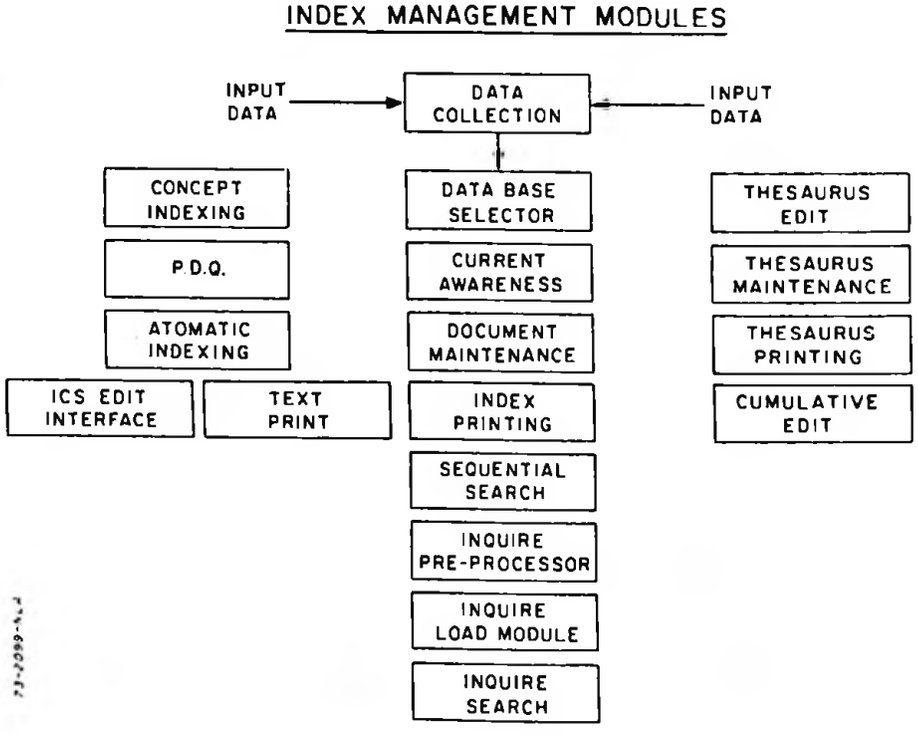

FIG. 1

We do not propose to discuss these modules in detail. However, we think these points are of interest: First, from a systems point of view, they can be regarded as standing alone. Suppose, for instance, some new development in either hardware or software suggests we might wish to make a change to the Document Maintenance Module. We can make it without touching any of the other modules.

Second, from a user-point of view, we can play all kinds of games. For instance, for any given information bank, we can elect to use concept indexing, P.D.Q. (our acronym for KWIC indexing) or automatic indexing, or we can combine types of indexing - we might, for instance, elect to put some deliberately selected concepts into an automatic indexing effort. We can build a thesaurus or by-pass that option. In fact, our options are almost limitless. However, once we choose a "path" through the modules for a particular task, the system reacts to it as if it was a single program.

The introduction of this degree of modularity into the system has had two major effects on our operations. First, it made it possible for us to develop our Microform Information Transfer Center, and second, it significantly lowered our operating costs.

In the M.I.T.C., which is shown schematically in Figure 2, COM-produced indexes are searched in a 16 m.m. cartridge reader-printer, and search results are retrieved in full-text from a microfiche file which is an integral part of the center. Because it is possible to perform Boolean searches of the indexes, and because search results can be delivered to the client as full-text, hard copy, with the help of the 16 m.m. fiche reader-printer, we occasionally and not entirely facetiously refer to it as our real-time, on-line, full text IR system. Its successful operation owes a great deal to our information bank philosophy and to the user being able to recognize which information bank is likely to contain the information he needs.

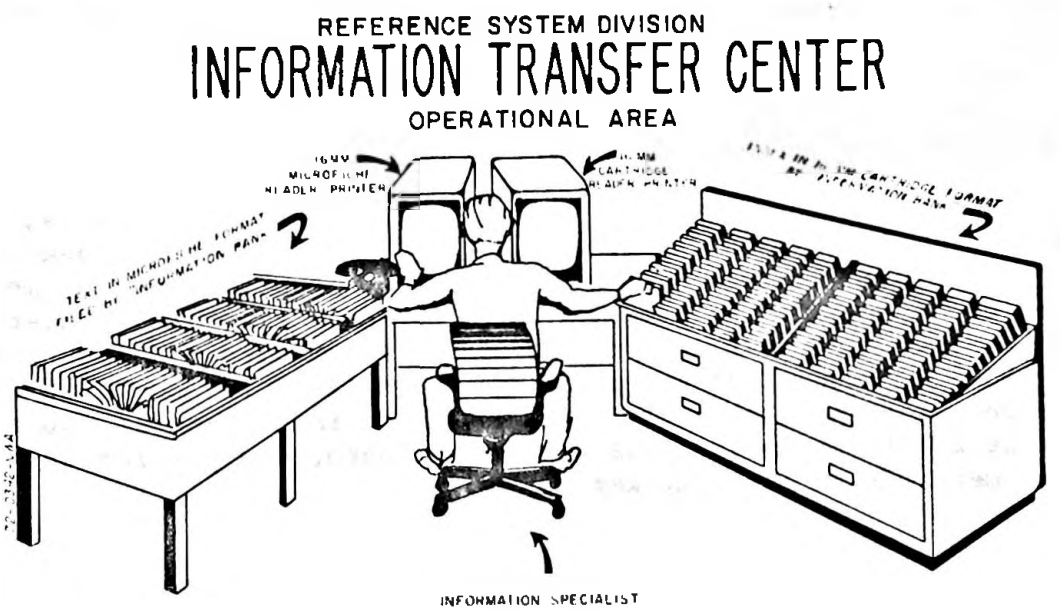

FIG. 2

Two comments before we leave the M.I.T.C. topic. One: - a COM-produced index costs 1/10th of its paper equivalent and is much easier to search. Second: - we can advance to real-time search on the computer as soon as demand for information and/or number of documents in any particular information bank indicate we can economically justify it. We merely begin to use the INQUIRE modules.

Now a word about costs. Operating with the system, it costs between 20 cents and $1.00 per document to update an information bank. The bigger the bank, i.e., the more documents it controls, the lower the cost per document. (How long this will continue to be true, we are not prepared to estimate.)

However, this is not too meaningful a figure, even if I tell you our computer charges are based on a cost of ± $800 per hour's CPU time, so let me put it this way. Operating with the IMM, we very quickly found that it cost us appreciably more for key-punching to get information into the computer than it did to manipulate that information once we'd got it in. It was this horrible fact of life that generated our interest in micro-computers - devices that offered the possibility of removing key-punch right out of our operation. Let me hasten to add that we quickly discovered they had much more to offer.

We make a subtle difference between micro-computers and mini-computers. A micro-computer can be called and regarded as an "intelligent terminal". It "stands alone" until the moment you want it to communicate with another and generally larger computer. The one we use has a built-in 12K memory (which is more memory than the second generation 1401's had).

Generally, it uses two cassettes, one to operate the programs we have developed for it and one to hold the information we are feeding into it. At any point of time, we can transmit the information on the second cassette, by telephone line, to our main computer some twenty-five miles away from our operating centre. When the main computer, a 370/165, has done whatever it is asked to do - it can transmit the results back to us and put the micro-computer back into business. We will return to that thesis in a moment but first of all we'd like to present some figures which we think might interest you.

At this moment, we have thirteen active Information Banks, and they control just over 27,000 documents. A recent analysis suggested that to get this control, we had to generate on average and let me emphasize "on average" 13.62 punch cards per document. The mini-computer we are using, with its printer and telecommunication link rents for ± $350 per month. If key punching and verification costs 10 cents per card and we believe that's about the going rate, then if you are entering 260 documents into the system per month, you pay for this rental merely by eliminating key punch.

The advantages of using a micro-computer, however, go further than this. Our indexers no longer have to fill in a computer-input-document. Rather, they examine a document, decide what key words they are going to use to index it, make a few notes and then feed the necessary information into the micro-computer.

We ran a time-test on this process recently. It suggested that on average "note-preparation time" took 4.19 minutes per document and "entry-time" took 5.50 minutes per document. If you figure that in New York an indexer is paid, including overheads and fringe benefits, $8.25 per hour or $1386 per month, then this test suggests that it costs about $1.30 to index a document and feed it into the micro-computer, ready for transmission to the main computer.

Let me sound a note of caution. These figures are based on N.Y. rates, and the type of documents we are working with. They are, we repeat averages. For comparative purposes, the same test indicated that using our old method of filling in an indexing input form took 7.85 minutes per document and of course, the information on the input form had to be key-punched before we could get it into the computer.

Now somewhat quickly we'd like to tell you a little more about how we operate with our micro-computer. It's not much bigger than an IBM Selectric typewriter and its keyboard is almost exactly the same. The main difference is that it has a T.V. screen attached, and it looks like this. (Fig. 3)

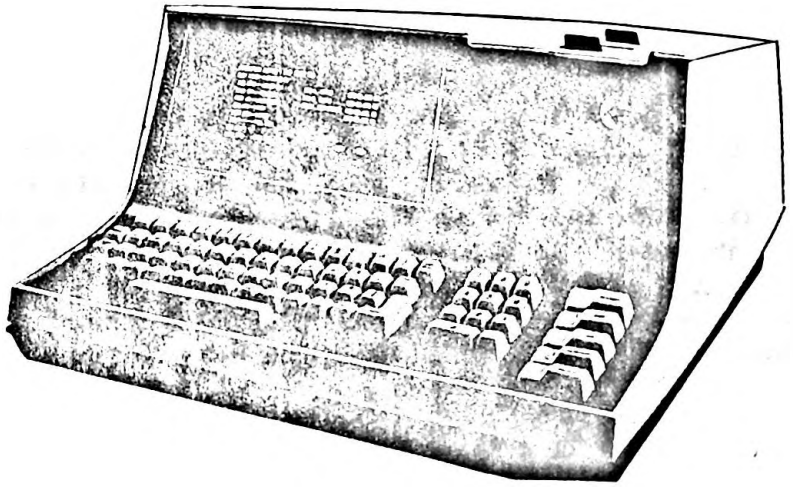

FIG. 3

Most people's first reaction to it is that the screen is very small. However, it can hold 960 characters - the same number as the very much larger IBM model 2260 CRT device.

The excellent character resolution and screen formatting capabilities built into soon remove this early impression.

We've set it up so that the user, having identified himself and been given certain pre-cautionary instructions, is shown what we call the "Menu" and it looks like this. (Fig. 4)

INDEX MANAGEMENT MODULE

1. CONCEPT INDEXING	6. RETRIEVE
2. DOCUMENT REVISIONS	7. PILOT
3. THES REVISIONS	8. D B UPDATE
4. P D Q	9. FILE SCAN
5. TEXT PROCESSING	10. TEXT EDIT

11. WRAP-UP

TO EXIT FROM ANY PROGRAM, PRESS & HOLD KEYBOARD
THEN PRESS ENTER KEY.

ENTER THE NUMBER CORRESPONDING TO THE
DESIRED FUNCTION .....ØØØ.

FIG. 4

In effect, the terminal is asking him what he wants to do. Now we would like to "for instance" it for you. Suppose he wishes to enter some Concept Indexing: he Keys-in 1 (at this point a cursor is blinking in the last position on the screen. .....ØØØ in Fig. 4). The terminal then searches for the appropriate program and in fact, shows him the equivalent of the Concept Indexing Input Form - line by line. When he has reached the 10th line, his screen looks like this. (Fig. 5)

CONCEPT INDEXING INPUT FORM

```
Ø1  D B
Ø2  DOCUMENT NO. ØØØØØØ
Ø3  D T C: ØØ
Ø4  LOGICAL SET OPTIONS:  (1Ø MAXIMUM).....
Ø5  DATE ØØ/ØØ/ØØ
Ø6  REFERENCE NO:
Ø7  TITLES 1 (T11):
Ø8           (T12):
Ø9  TITLE  2 (T21):
1Ø           (T22):

ARE THERE ANY CHANGES:  ( Y OR N )
```

FIG. 5

When he has indicated that he does not wish to make any further changes, the screen format changes and he is shown how he should enter his keywords, get them into his thesaurus and "weight" them. Note the data is edited as it is entered and provision is made for making revisions to input data.

Obviously, time does not allow us to discuss our other options. We incline to the belief that you'ld like them. So let's leave this aspect of our micro-computer and just indicate one other thing about it which excites us.

We have found, to our great delight that it is a most efficient means of access to remote "public utility" information banks. For instance, we can access Lockheed's information bank in California and query the ERIC data base, the NTIS data base, and/or the Pandex data base at a price we can afford, i.e., $25 to $35 per terminal connect hour. Because we provide our own terminal and communications (through our tie-line system) this is the only charge we incur. Remember too, that the elimination of key-punch has paid for the micro-computer.