create paper in such a way, and they indeed would be responsible for filing it in the first place, so that a single piece of paper would belong in one and only one cubby hole. Information and paper generated referring to many separate classifications should not exist, providing the user co-operates.

A. S. Rettig (Radio Corporation of America): How do you propose to protect any of the documents filed so that, if removed, they may be retrieved?

Dr. Fein: This is a problem of control of documents that most libraries have.

The way it was done here was to have a custodian of the files who actually checked out the information removed, similar to that which a librarian uses when you check out a book in a local public library. The effectiveness of the control procedure depends upon the co-operation of the user.

File Problems Associated with The National Menu Study

P. M. THOMPSON

THE ORGANIZATION and structure of an information file cannot properly be discussed by itself. The design of a file only makes sense when it is considered as part of the entire problem of information retrieval and maintenance of the file. In any particular instance consideration must be given to the expected frequency of use of the file, and to the particular machine logic that will be employed both in retrieval and in maintenance. A properly designed information file will reflect an optimization of some sort, weighting among other things speed, accuracy, and cost. The filing and retrieval activity, of course, is usually only a part of a total information processing system. A balanced over-all design of the entire machine system must, in turn, be achieved which will include not only all aspects of the information processing at the electronic center but in addition, all available criteria by which one can discover the proper relationships among the electronic data-processing system and the basic objectives of the entire business enterprise.

The Menu Study information file is unique from several points of view. It is not part of a routine data-processing system, but is of a one-shot nature. This file of information resulted from a very great field study of American eating habits. It is being subjected to many specific analyses by a number of food companies seeking information in very much the same way that they would were they to conduct a series of (relatively) small field surveys. It was the aim of the menu study to put enough data into one large file so that “field studies” could be made at the machine rather than in the field. There are many advantages to this approach once sufficient financial backing is available to make it possible. Relatively short questions can be formulated and answered by machine quickly and inexpensively. Research questions characteristically lead sequentially from one to another. With this large study, the data on tape are sufficient to answer large quantities of question sequences for a great variety of food interests. Thus it is seen that many and varied questions will be put to these data.

The National Menu Study

The National Menu Study was undertaken by the Market Research Corporation of America with underwriting cooperation of several large food processing companies. A representative sample of about 2,000 families recorded the details of well over one million menu item servings. About 2,000 menu item choices are defined in a 5-level “dish” code structure. The ingredients of all home cooked dishes are coded and included in the file. In addition to classification codes for the individual dishes and ingredients, more than 50 other measurements were taken. They include other classification data related to the serving of the dish, and also classification data describing the meal and day of the serving and of the family serving. Examples of the measurements taken are:

- Family: Geographic location, age, sex, and diet status of individual members, family income level, education of head, etc.
- Day: Weather, who shopped for food, number of hours the housewife was away from home, etc.
- Meal: Type of meal, where eaten, time, family members and guests present at meal, etc.
- Dish: Family members and guests eating the dish, who prepared it, new or left-over, fundamental dish or additive, etc.

It is particularly important to note that all information is coded into discrete cells. A family with an income falling between $6,000 and $6,999, for example, is coded 6 under “income.” A family member 13 to 17 years old would be coded 3 under “age.” Thus, nowhere in the file is the actual value of a measurement recorded, only coded representations with each code number covering a cell. In general, the entire range of each variable is described in 10 or fewer cells. The greatest exceptions are the first and second levels of dish codes where 64 cells are available for each level.

Typical Analyses

In a typical analysis of these data, only two kinds of basic operations occur: 1. A selection of an element of data is made, and 2. some counting function is applied to the selected elements. For example, a single question to ask would be: How many families served lobster, how many servings were there, how many people were present when lobster was served, and how many people ate lobster? In this example the selection operation involved selecting out of the data all servings of lobster. Under the first level of dish coding, code number 23 covers fish and shell fish. Under first level 23, the second level code number, 57, covers lobster. So the selection function is simply the logical expression (first level 23 and second level 57). Whenever this logical product is equal to one, the dish is lobster. If the product is zero, the dish is not lobster.

The counting function is applied only when the selection function is satisfied (i.e., \( F = 1 \)). In addition to the counts indicated, various distributions may be indicated. For example, one may wish to know the number of families who serve frozen beef dinners (a convenience item) on days the housewife is away from home 6 or more hours, and the distribution of those families over, say, income. The select function in this case would be: First level dish code 62 (commercial pies and prepared dinners) and second level code 31 (beef dinners) and fifth level code

P. M. THOMPSON is with the Market Research Corporation of America, Chicago, Ill.
Data Organization

All of the data have been reduced to binary form and packed as tightly as possible on tapes for use with an International Business Machines Corporation 704 or 709. Since families are individually and randomly selected for the sample, there are no functional relationships between families. Thus a select function need never encompass the data of more than one family at a time. One record of information has been created for each family and these records are brought into memory one at a time. All analyses are completed for one family at a time with no loss of generality.

For a given family, the data fall into a true hierarchical structure. Ingredients are inferior to dishes which are inferior to meals which are inferior to days which, in turn, are inferior to a given family. Advantage is taken of this structure to allow penetration into lower levels of data only when a select function is satisfied at higher levels. This saves much time in looking at data. For instance, if a question is asked for steak consumption by low income-level families, the dish level would not be examined for those families that did not qualify on income.

The family record is a variable length record since the number of ingredients, dishes, and meals vary from one family to the next. There was even little variation in the number of days each family reported. The resulting records vary in length from several hundred 704 words to about 2,500.

The family data record is composed as follows: The first-12 words contain family classification data. The next group of words contains day classification data with two words required for each day the family reported. The next group of words contains meal classification data with two words required for each meal. Next are the dish words with two words per dish, and finally the ingredient words where each word contains three ingredients.

In order for the machine code to penetrate the family record, the following address information is contained in the record. The number of day words is contained at a known location in the family classification words at the beginning of the record. Since the first day word is the 13th word of the record, the data for any given day can be retrieved with a single address computation. In each pair of “day words” the location and number of meal words for that day are given. Thus, any given meal can be retrieved with two levels of address computation. Similarly, each meal group contains the location and number of dish words for that meal, and each dish group contains the number and location of the ingredients words for that dish plus the number of ingredients in the dish. Thus, with a maximum of four levels of address computation, any given element of data can be retrieved. If it is desired to simply examine all meal information, the location and number of all meal words are contained in the family words. The same is true for the dish words. The location and number of dish applied to an index register loop will examine all of the dish words directly.

Requirements for the Machine Code

The Menu Data is being subjected to a large number of fairly restrictive questions. Complex analyses are composed of sequential questions with successive questions reflecting knowledge gained in previous questions. An automatic machine compiler is required to prepare the individual machine codes for the various questions. To program individual questions would be too slow and costly. The code must work with variable length records. It must search only through pertinent data in the hierarchical levels of the family record. The compiler must allow matching of enough questions to fully utilize the machine memory during a pass.

There are two working codes in memory at object time. One code performs the selecting functions and the other the counting functions. The select code is simply a code that will evaluate any logical function that may be encountered. Whenever the value of the select function is unity, the select code transfers control to the count code. The count code is essentially a collection of individual routines for the various count functions.

If secondary selection is required (within primary selection), the count code, which was given control by the primary select function, passes control back to the select code after it completes the necessary counting for the primary selection. It passes along the parameters to define the secondary selection. Meanwhile the primary select parameters are preserved along with the location in the data record where the current primary selection occurred. When the secondary selection and its associated counting is completed, control will be passed back to continue the primary selection. This cyclical process is continued until the family record has been completely examined by the primary select function. The process is then repeated for succes-
sive questions until all questions in a batch have been completed. At this time a new family record is read and the entire process repeated.

The select code will consider as a unit of operation the following:

1. It will examine a logical product, or a logical sum, of as many terms as required with the extension that any number of specific codes can be tested for any variable of measurement on a logical or basis. That is to say, if one term in the logical product or sum involves say, income, several income cells can be treated on an or basis (i.e., income cell 5 or 6 or 7).

2. It will evaluate this sum or product for one combination of data involving a specific day, meal, and dish for the family who is in memory. The data are examined in descending hierarchical order and as soon as a logical product is determined to be zero or a sum to be unity the work is stopped, for the final value is by then determined.

More complex select functions are built up with successive levels of parentheses of sums or products. The inner sums or products are evaluated first (found to be zero or unity) whereupon they are then treated as single terms at the next higher level.

Parameters and other operating information are supplied to both the select and the count codes by means of tables. Each table may be considered to be a calling sequence though the codes are not employed in the usual subroutine sense. A table will supply to the select code the form of the logical expression, the specific code values that make up its terms, and the masks necessary for unpacking. In addition it will give the location of the associated counting code tables that are to be used. The counting tables will provide information as to the counting functions applicable and will give the location of any secondary select tables that may be required. They in turn give the location of secondary counting tables. Thus a chain is formed. Control will be passed as far down the chain as required and will return the same chain to the primary select table which supplies the master control for advancing through the family record.

The Compiler

The compiler has a relatively simple task. Information is presented to the compiler in a language centered about Menu Study nomenclature. The compiler has access to a large reference table where it can obtain specific address information, masks, etc. The compiler translates the questions from the menu language into the necessary sets of "calling sequence" tables. Output data space is an integral part of the count code tables. Printout heading information can be associated with the output data space for identification or results. The amount of table space required per question can quickly be determined and thus the number of questions that can be handled per batch established. The space requirements for the family record and for the select and count codes are known. Allocation of the remaining space is simply a matter of putting in as many tables as the space will hold.

All of the counting tables are located together in one section of memory. With the assistance of incremental address information available in the counting tables, a printing routine can work its way through the locations containing output data and associated headings.

Discussion

E. Herscher (Philco Corporation): When searching for the dishes that were served with steak, how does the system determine the dishes, in the same meal, that it passed before reaching steak?

Mr. Thompson: The steak in primary selection, and one set of index registers keeps track of where we are in the record on this selection. When control is passed through the accounting code back to the selecting code, another set of index registers takes over the secondary selectors, and the boundaries are set up by the mode of association, and it is completely independent. When this selection is completed, control is passed to the primary register and the original index registers pick up where they left off.

Data Processing and Information Handling

R. H. GREGORY M. TRUST

It is generally recognized that, at the present stage of development, businesses must process data to produce reports useful for management guidance in making decisions. Such reports are, however, often more dependent upon what data are available and the mechanics of processing than upon managerial needs for facts and abilities to use them. This arrangement might be called the "push" or "supply" approach to data processing.

An alternate arrangement to producing reports required for managerial purposes might be called the "pull" or "demand" approach. Reports required for managerial action are explicitly specified and a system, including both processing and data origination, is devised to produce the desired results. These two arrangements for getting useful reports from raw facts are extreme cases and many intermediate schemes exist.

Data and Information

In dealing with managerial reporting it is useful to draw a distinction between all of the facts available, "data," and those used for decision making, "information."

Data

Data can be defined as any facts that are a matter of direct observation, are known or available, and may be expressed as numbers, words, charts, or tables. Raw data arising from business transactions can be processed in the "push" or "supply" fashion, as described, to yield files and reports that might be called "processed data." Raw and processed data are interesting but may not be useful for management decision making.

Information

Information can be used to mean data, either raw or processed, that are new, accurate, and timely. A manager obtains information from a report if he learns something he did not know before, if the facts are accurate enough for the situation involved, and if the report is obtained in time for him to take action. Reports are

Gregory, Trust—Data Processing and Information Handling 65