GIM-1, a generalized information management language and computer system

by DONALD B. NELSON, RICHARD A. PICK
AND KENTON B. ANDREWS

TRW Systems
Redondo Beach, California

The current demand for more comprehensive information systems had created an increasing need for generalized information management. Generalized information management denotes a complex of interrelated information capabilities encompassing both human functions and the operation of automatic devices which supplement and extend human capabilities. The methodology and techniques for such generalized information control have been defined and are exemplified in GIM-1, a generalized information management language and computer system now being implemented at TRW Systems. With the use of GIM-1, the problems inherent in defining more comprehensive information systems can be completely and economically resolved.

The definition of a system for such information management requires a communication network of many remote stations and a central complex of one or more computers. It also must accommodate natural language queries and multiple technical vocabularies. Additionally, such a system must provide automatic correlation of information both within and between data files, complete as well as selectively limited information security, and automatic controls for data reliability, data conversions and the synthesis of data as a function of other data. All of these requirements for comprehensive information transfer can be satisfied by the use of GIM-1.

GIM-1 permits natural language access to data list information stored in a random bulk storage device which also is available to other computer programs, and the remote user may select any available equipment in the entire network for an information output. Although GIM-1 is general to information management needs, its design accommodates the particular information requirements of any one application area with special purpose efficiency in computer operating times. GIM-1 is essentially machine independent, and the calculated addressing scheme permits retrieval of any required data from random storage in a single access.

The GIM-1 language is limited natural English, and the formats and rules for remote inputs are both simple and very general. The language processors, together with the use of dictionaries, permit inputs to be stated directly in the technical terminology natural to each application area, and also provides the user with plain language outputs. The GIM-1 language uses the lineal format natural to prose text, accepts any number of variable length words, and permits a limited freedom of word order. Minimal vocabulary prohibitions to ensure uniqueness apply only to the data list identifiers and are guaranteed by automatic language audits, but no vocabulary prohibitions are imposed on the information values. Additionally, the GIM-1 language permits the definition of any number of identifier synonyms, and provides a large selection of input conditionals which may be used to limit outputs to only relevant information.

In GIM-1, the “master” dictionary contains the data list identifier nouns, and the “user” dictionary associated with each data list contains the attribute identifier nouns. The master dictionary also contains the process identifier verbs and the language connectives; and, in each of these dictionaries, any number of synonyms may be defined for each identifier. The connective words include all conjunctions, prepositions, articles and special symbols specified for possible use in language inputs. The definition
of these connectives, then, includes the designation of relational operators and other conditional words for limiting information searches to only relevant data. Therefore, the specification of connectives will dictate operational procedures not only for the language preprocessor but also for the automatic correlation of data and the numerous data processors within GIM-1. Additionally, a limited natural syntax has been defined for GIM-1, since any completely natural syntax would require a probabilistic processor contradictory both to practical efficiency and to the absolute identifications required for updating data lists with complete accuracy.

The GIM-1 data base consists of many different data lists, and all GIM-1 information is stored in data lists. Each data list is identified by a DATA LIST I. D. and consists of many different items of information; and each item consists of an ITEM I. D. followed or not by relevant information values. All information values, then, are stored in items within data lists, and each VALUE is identified by one of the many attributes defined for each data list. Each attribute is identified by an ATTRIBUTE I. D. and any number of attributes can be defined for the identification of all possible data list values. However, since items contain only relevant values or none at all, the usual document has values for only a few of the many attributes defined for a data list.

The standard format for organizing GIM-1 information, then, has a column heading for each of the following four information elements:

<table>
<thead>
<tr>
<th>Information element</th>
<th>Mnemonic Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA LIST I.D.</td>
<td>D</td>
</tr>
<tr>
<td>ITEM I.D.</td>
<td>I</td>
</tr>
<tr>
<td>ATTRIBUTE I.D.</td>
<td>A</td>
</tr>
<tr>
<td>VALUE</td>
<td>V</td>
</tr>
</tbody>
</table>

An example of information organized in accordance with such a format is illustrated in Figure 1.

Since a data list may have virtually unlimited attributes which may be used or not in any given item, a user is neither restricted to a small set of fixed identifiers nor required to assign a value to each identifier for every item. For example, assume all references for transportation systems are itemized by library code, with each reference containing such standard information as TITLE, AUTHOR, ABSTRACT, etc. In addition to these standard attributes, however, much greater indexing depth can be achieved by also defining many special attributes such as GROUND EFFECT MACHINES, HELICOPTERS, MONORAILS, TOTAL COST, and PRINCIPAL CITY. These extra attributes, then, define special information which may or may not be relevant to any one transportation system reference. For instance, the majority of references may include absolutely no information for these extra attributes and, therefore, no extra values would be stored. However, for any item which may contain such special information, these extra attributes would enable an indexer to describe the reference item in much greater depth. Provided such attributes were defined, then, a GIM-1 user might initiate the following input:

```
LIST THE TITLE AUTHOR AND ABSTRACT OF EVERY TRANSPORTATION SYSTEM REFERENCE WITH THE PRINCIPAL CITY "LOS ANGELES" AND THE GROUND EFFECT MACHINES "AVC-1" OR "HOVERCRAFT"
```

Furthermore, the organization of GIM-1 information permits the remote user to add new attributes for a data list whenever he may wish, and without any effect on the existing data list information. For instance, an indexer might encounter a transportation system reference which included a description of the net change in smog levels during trial tests of some proposed system. Assuming all of the existing attributes to be unsuitable for a description of this information, the indexer might wish to add the new attribute HYDROCARBONS. This new attribute then would be available for possible use in any old or new transportation system reference.

The four elements within the standard format for GIM-1 information can be considered both as an attribute value DATUM having three IDENTIFIERS (Figure 2), and as a set of two dyads, with each dyad having one DATUM and one IDENTIFIER (Figure 3).

Consideration of the standard format as a sequence of two dyads of information is a concept of particular importance. Together with the automatic correlation of GIM-1 data, both within and between data lists, this concept permits the definition of extensive and very complex information formats. The construction of an extended data format by a unidirectional chaining of dyad units is illustrated in Figure 4.

As an elementary example of data format extension, assume the unidirectional construction in Figure 5 as defining both the retrieval and update-add correlation between the attribute AUTHOR in the ANEW DOCUMENT data list (defined in Figure 1) and another data list identified by TECH/AUTHOR.

In Figure 6, therefore, any value for AUTHOR automatically will be also an item I. D. for TECH/
The elementary correlations in Figures 4, 5, and 6 are constructed by unidirectional chaining of the values of an attribute in one DIAV unit with the item I.D. in another DIAV unit. However, many other types of automatic correlation also are available to the GIM-1 user. For instance, “Christmas Tree” interrelationships between different items within the same data list require bidirectional chaining of the values of an attribute with the item I.D. in a single DIAV unit. Additionally, some commonly required correlations are considerably more complex. For instance many interrelationships require correlations between the values of two attributes in different data lists, together with automatic item identifications. This type of correlation is illustrated in Figure 7 as an elementary construction with only two unidirectional correlatives and with only one correlative defined for the values of an attribute.

In Figure 7 then, each update of a value in D1A3 requires the automatic update of an item I.D. in D2. Similarly, each update of a value in D1A4 requires the automatic update of a value in D2A4. However, the automatic update of D2A4 further requires the automatic identification of each item I.D. in D2 which also is defined as a value in D1A3. The automatic creation of such “bridge” values
between data lists, then, is required for any correlation between the values of two attributes in different data lists. Although these few examples of automatic correlation are elementary, they are sufficient to indicate the flexibility provided the GIM-I user in defining extensions of the standard format for organizing information.

GIM-I, then, permits the user to specify very complex data list interrelationships for automatic correlation. Only a simple dictionary code enforces any defined correlation for all subsequent inputs, and GIM-I processes all interrelated information with total relevance and with complete accuracy. The several correlations available for specification by the GIM-I user include:

- The automatic retrieval of associated information stored in other data lists.
- The automatic updating of correlated information stored in other data lists.
- The automatic updating of cross-indexed data lists.
- The automatic calculation of data values as a function of other stored values.
- The automatic coupling of unit values in associated sequences of multiple values.
- The automatic creation of “bridge” values for correlation between data lists and items.
- The automatic updating of all relevant information in “Christmas tree” data lists.
- Full “Christmas tree” searches for retrieval and counting requests.
- The automatic retrieval and updating of non-redundant information stored only in another data list.
Among these several optional capabilities defined for user convenience, the automatic updating of information in correlated items and data lists is of particular importance.

For example, a user might require optional access to information by classifications unique to his own technical area, while the technical library might wish to file new information only by DDC index number. These apparently different user needs are not contradictory, however, since both requirements can be accommodated by the automatic updating of correlated information. Assuming the information for each DDC index number is defined to include such attribute identifiers as TITLE, AUTHOR, ABSTRACT, SOURCE, TECHNICAL AREA and DESCRIPTORS, and also assuming the relevant correlation codes are stored in the dictionaries, the GIM-I system would update not only the DDC data list stated in the technical library input, but also automatically would update any number of correlated data lists and items defined by the dictionary codes and the values stored under the encoded identifiers. This complex capability, therefore, eliminates not only the need for complex intervention by the user but also the probability of human error, and ensures the complete accuracy and relevance of interrelated information.

GIM-I further accommodates the remote user by providing complete as well as selectively limited information security, automatic value conversions and complex format audits of input values. Additionally, the executive control system provides priority processing and automatic record-keeping of all transactions. The standard data processing capabilities of GIM-I include the retrieval of stored information; the counting of information values; the deletion, addition or changing of stored information; the remote initiation of new dictionaries and data lists and a special report processor for generating special output formats.

As exemplified by GIM-I, then, generalized information management is based on a new and interdisciplinary approach to the total problem of information, and is a comprehensive complex of interrelated capabilities. As indicated by its name, however, GIM-I is only an initial competence with many extensions of present capabilities, as well as new capabilities, planned for the future. At present, for example, the automatic updating of information between data lists provides a selective dissemination of information, but only within the GIM-I data bank. However, the rapidly growing need for selective dissemination requires an external distribution; and, because of the construction of functional group interest profiles, the impetus for the flow of information rests with the system rather than with the user. Therefore, in a future version of GIM-I, the present selective dissemination of information will be extended to include automatic decision-information outputs to all users. Similarly, the increasing demand for international information transfer requires a system which accommodates multiple languages, as well as multiple technical vocabularies, with direct access to mutually shared data banks without language translation. However, GIM-I already accommodates multiple technical vocabularies with direct access to mutually shared data lists without vocabulary translation. Additionally, GIM-I is both machine independent and language dependent. Therefore, in a future version of GIM-I, the present natural English language will be supplemented by Russian, French and German.

This description of GIM-I has been brief, and the summary of its capabilities gross. As an example of generalized information management, however, it has been sufficient to indicate the methodology and some of the techniques for resolving the problems inherent in designing more comprehensive information systems. Also, it has indicated the need for integrating the definitions of company information standards and management responsibilities with the specification of a communication network and the software control system. And hopefully, this brief description of GIM-I has been sufficient to verify generalized information management as a new and powerful tool for supplementing and extending present capabilities in information systems.