

Construction Database Management by Natural Language

ROOZBEH KANGARI

The objective of this paper is to show how contractors can implement microcomputers to database management systems by natural language to easily analyze the mountain of information generated in today's construction environment. Conventional database management systems require that the person using a microcomputer have extensive knowledge of the contents and organization of the database. The natural language system allows an inexperienced contractor to retrieve information by asking questions in a conversational English manner, permitting the production of reports without the intervention of the data processing department. Two commercially available systems designed for microcomputers are reviewed, and their application in construction is discussed.

There are growing emphases on making microcomputers more user-friendly to interact with human beings without the people's needing to be computer specialists. The main purpose of a natural language system is to hide from the user the technical details of the system with which the user is communicating, for example, the way in which a database is structured. In other words, it transforms the request as expressed by a user, in a form that the user finds natural, into the formal expression required by the computer.

Database management has been an early target of the natural language community. Many natural language systems have been developed to ease the interaction burden between the user and the information management system. The most important aspect of a natural language system from a contractor's viewpoint is its potential for helping engineers and managers obtain information about a construction project from a database without the aid of a computer professional.

The term database describes a collection of information. For example, name, address, and telephone number of material suppliers are a database, or construction cost data are a database. In other words, a database is an electronic file cabinet, containing information belonging to a company. Subject categories in the database are called files, such as customers, employees, vendors, and the like. Information about an individual member within a category is called a record. Items within a record, such as an employee's social security number or a vendor's address, are called fields (1).

Using natural language, a construction manager can ask the question in a conversational way without having to think

about the way in which the question should be constructed. It is much easier to type, "Show me the average monthly cost of labor for project 372," than to list all the fields in the cost file that should be checked and give the computer a formula for calculating an average. People in the construction industry who don't know the computer's complex language, and have no time to learn it, are demanding access to computer data storage and retrieval capabilities. That means that more efficient natural language systems will be developed to understand the user's natural language rather than requiring the user to learn the computer's language.

NATURAL LANGUAGE SYSTEMS

Constructing a natural language interface between human user and computer is a problem on which one branch of the artificial intelligence community has been working since the 1960s. In the beginning, experimental natural language systems were built to contain information about a tightly limited domain. The system scanned input statements, searching for key words or predefined patterns. These inputs, when they were found, brought into play rules for responding to the input with patterns of apparently appropriate words from the system's database. The ambiguities of language and the sheer size of the vocabularies of highly developed languages remain formidable obstacles to completely flexible interactive dialog. Yet recently, a number of microcomputer-based natural language systems with excellent performance have entered the marketplace (2).

Any natural language system must contain a list of words that the program will use in preparing its reply. Natural language systems also contain a syntax that specifies the order in which words must appear. Also, the system contains a set of transformation rules that determine how user input will be converted into system output.

In order to answer a database query, a natural language system performs the following tasks:

1. Identifies specific fields of the database to be inspected,
2. Identifies records in the database whose entries in the specified fields match the stated criteria,
3. Determines the retrieval and display processes required to respond to the request, and
4. Invokes these processes in the proper sequence.

No computer system can possibly anticipate all of the questions that can be asked about a specified subject. The

problem is compounded on a microcomputer, even on one equipped with a hard disk. No computer or software has yet been built that is powerful enough to match the spontaneous exchange of ideas that occurs naturally between human beings. But, still within its limitations, adding a natural language system to a database makes it accessible to industries such as construction that most need to use it (3).

One of the most important and feasible areas for the application of a natural language system in construction management is in accessing data in databases. Thousands of dollars are usually spent in large construction projects in collecting data. However, this information is generally not readily available to the people who need it. For example, a project manager wants to direct a simple question to a microcomputer. The project manager wants to know, "What activity in project No. 274 has the highest cost overrun?" The project manager knows the information is in the PC, but lacks the expertise of knowing the computer language. The project manager must find a computer programmer who can translate his question into a formal query to give to the PC. Unfortunately, the programmer is working on an important project. Even when a programmer is available, misunderstandings often occur and there are problems in creating proper code. By the time an answer is extracted from the computer, it may no longer be timely and may not even be relevant.

MICROCOMPUTER TOOLS

There are many reasons that make microcomputer-based natural language systems attractive to the construction industry. First, the managers of construction projects are not generally familiar with the details of the structure of the database. Second, these nonexpert people forget the complex rules of computer language; and finally, most people involved in construction are not computer specialists and have little enthusiasm for learning a computer language with its rigid syntax. Therefore, natural language processing systems can ease the process of getting information and improve access to a company's database.

Capabilities of two commercially available microcomputer-based natural language systems for database management are presented. These two programs are (a) Clout2 and (b) Savvy.

Clout2 is a database management system by natural language that allows users to ask questions in conversational English. It can be implemented by PC programs like Lotus 1-2-3, PFS:FILE, dBASE II, Multiplan, and others. It is equipped with a 300-word vocabulary, and the user can add up to 500 additional words for each database. It has automatic data scanning and cross referencing, so there is no need to remember how data are stored. Clout2 learns new phrases in a dialog manner. New words or phrases can be

added to the permanent dictionary. The system has an automatic spelling checker and on-line help. For example, a project manager can get clear and quick answers by asking questions such as, "What is the average cost of 24 inches of plywood column?" or "Smith Construction Material Supplier Phone Number?" If a term is used that the system cannot understand, it initiates a dialog designed to build a definition for the term.

Savvy is another database system by natural language that uses pattern recognition algorithms. There are two ways to create a database: with the programming language, which is English-like in its syntax and can be used by nonprogrammers; and by using an automatic programmer, which strings together appropriate commands after the user has designed forms on which to collect and store data.

Implementation of these systems is an important step toward the practical use of microcomputers in the construction industry. Contractors with no experience can now use plain English to create database files, sort cost items, and manipulate and retrieve information in free-form plain language style. Construction project engineers are no longer confined to a system of codes and symbols that can make programming cumbersome and time-consuming.

SUMMARY AND CONCLUSIONS

Database management by natural language in a microcomputer is an excellent system allowing contractors to interact naturally with large and complex databases containing time and cost data, productivity information, accounting data, and other types of useful information. Although certainly far from perfect, many of these systems permit contractors to communicate with a computer in much the same way they might address a colleague, owner, or material supplier. For the first time, these project managers averse to or incapable of mastering complicated computer languages can access a database with little effort. Microcomputer implementation of these limited language processing techniques is leading to more practical and cost-effective systems.

REFERENCES

1. G. L. Simons. *Introducing Artificial Intelligence*. The National Computing Center Publications, Manchester, U.K., pp. 147-170, 1984.
2. D. A. Waterman. *A Guide to Expert Systems*. Addison-Wesley Publications, Reading, Mass., 1985.
3. S. J. Andriole. *Applications in Artificial Intelligence*. Petrocelli Books, Princeton, N.J., 1985, pp. 143-227.

Publication of this paper sponsored by Committee on Construction Management.