An Intelligent Retrieval and Presentation System for Management Information

M. RAMSHAW†, A. P. STEWARD†* AND D. STOVES‡

† *Department of Mathematics and Computer Science, Sunderland Polytechnic, Priestman Building, Green Terrace, Sunderland SR1 3SD*
‡ *British Gas Engineering Research Station, Newcastle upon Tyne*

A prototype for a system for free format access to management information has been developed jointly by British Gas and Sunderland Polytechnic. The design criteria and operational results are described.

1. INTRODUCTION

Like most large organisations British Gas has a wealth of information stored on computers. Access to this information is most frequently provided by printed reports available in varying degrees of detail. Some of this information can be accessed directly on VDUs using standard reporting formats, however access methods such as command languages create difficulties for the infrequent or non-expert user.

Senior management would benefit from access to information via a system with a simple interface which avoided the use of command languages and numerous interactions such as with typical menu-driven systems. Hence a joint project has been established between British Gas and Sunderland Polytechnique to investigate the feasibility of free-format access to management information. The major objectives of this project are to explore this possibility and the limitations of supplying replies to users' queries within a limited domain without: (i) a command language; (ii) the use of menus; (iii) the need to use several interactions; (iv) true natural-language understanding.

Unlike front ends to current database management systems this project is to provide an interface between the user's natural-language question and existing data files, where no DBMS is present. Thus it is necessary for our interface to contain the ability to: (i) interpret the user's question; (ii) determine the data required; (iii) retrieve the data; (iv) manipulate the base data into a form suitable to answer the user's question.

Further criteria were that the system must be capable of running on a PC and must be quick and simple to use, requiring no computer skills other than the ability to use a keyboard.

2. THE BASIS FOR THE SYSTEM DESIGN

One particular set of data which is of interest throughout British Gas, related to management information in engineering, takes the form of a structured set of tables, consisting of approximately sixty tables for each year.

Our overall aim was to produce an enquiry system based on this collection of tables which would allow the manager to input a request for data in free text and to have the request answered by appropriate output in a high proportion of instances. In cases when the system

was unable to interpret the request we proposed that the area of the data relevant to the request should be determined as closely as possible by inference and the user assisted in refining the request.

Each line of a table has been associated with a description presented in the printed tables. However, the form of words that is used to describe that activity by an individual when consulting the system in free format may well differ from the description in the tables. Equally, the construction of the request may vary from a list of words which the user imagines is sufficient to convey a request to a completely formed sentence. To accommodate the varying forms of request we have chosen to adopt a keyword and synonyms strategy which allows the request to have a free-text form.

In this approach each request is first scanned for words to be replaced by keywords and then the resulting set of keywords is used to match with a set of identifiers for each line of the tables. Also, to keep the interaction as elementary as possible we require that the request is typed as a whole and that the enquiry is processed in one stage.

Our questioning of users suggests that the kind of data which in practice they require are as follows.

(i) Statistics from a table on a given activity in a selection of, or all British Gas regions; giving where requested trend and averages. Should the activity requested be a subtotal or total line in the tables, then information about the components of the activity would be required.

(ii) Statistics which are not in the tables, but which can be calculated, in a selected region or all regions of the company, plus information showing the components of the calculation.

(iii) Presentation of further information about the components of an activity in question.

A computer language to implement this design would have to possess the following abilities: (i) to transform the free-text input into a set of words expressed as a list; (ii) to scan lists for words to be replaced; (iii) to manipulate lists and to test for sublists; (iv) to easily express the structure of the data. To satisfy these requirements and also in order that an element of parsing could be incorporated into future developments we decided to implement the prototype in Prolog. This language also has the advantage that extra facilities can be incorporated into the system without radical redesign of the program.

---

* To whom correspondence should be addressed.

3. THE PROTOTYPE

A prototype has been developed in Prolog on micro-computers using MS-DOS and PC-DOS. The data falls naturally into six main areas, and because of the limitations of Prolog the data has been divided into six parts, using this natural classification. Adopting this division of the data enables the system to read into memory only the data which is relevant to the enquiry; this obviously reduces the amount of memory used for data. The search space is also narrowed, giving a reduction in response time. However, using this classification of the data does incur a cost, as the system now needs a means of selecting which data area is required. A further reduction of memory requirement is achieved by reclaiming memory during execution by the use of controlled failures. The prototype is keyword-driven and the process depends upon the number of keywords or their synonyms which the user has included in the question.

These two requests produce the same response even though they are syntactically different.

> For Scottish Region, give me the number of new service pipes laid.

> How many new service connections have been installed in Scotland?

Response: the total number of new services laid in Scottish Region = 51.

4. DISCUSSION

A system based upon the design and solution strategies outlined has been developed, and works well when handling simple queries in a limited domain. It can currently access twelve tables of information, each containing approximately 200 individual pieces of information, covering a representative range of British Gas activities.

The prototype is designed to supply only primary statistics from the tables and perform limited operations, e.g. trend, but it is our intention to allow further operations on the data and that these operations are to be triggered by identifying such words in the query as 'compare' and 'greatest'.

The system has no memory for previous requests, so that a query must include sufficient information to specify which area of information is being requested. In future an element of recall of the immediately previous query will allow a user to make more efficient follow-up queries without repetition.

Although a specific set of data is being employed in this exercise it should be feasible to use these techniques to build a shell for more general data access.

This technique of providing 'natural language' query interpretation has potential to be employed on a mainframe as a front end to existing database query systems, where the need for command language is inhibiting and yet in other respects such as for statistical manipulation they are often very powerful. Already output from a DBMS on a mainframe has been handled by the prototype.

The system described is still being developed and the techniques are being explored. It is our intention that this will help us appreciate some of the possibilities and their limitations and that some useful tools may emerge from the exercise.