An expert system for upgrading small water supplies

LIBRARY
INTERNATIONAL REFERENCE CENTRE
FOR COMMUNITY WATER SUPPLY AND
SANITATION (IRC)

201
87EX

B. Knight

This paper describes an expert system developed by Thames Polytechnic in collaboration with the Water Research Centre (WRC) to assist in the selection of process units suitable for upgrading small water supplies. The objective of the development was to produce a system which could be evaluated more for its practical utility than for sophistication of its artificial intelligence (AI) concepts. This overall objective guided each phase of the development, from initial selection of the expert task through logical and physical design of the system to final implementation and review.

Keywords: expert systems, upgrading, water supplies, artificial intelligence

The task of upgrading small water supplies was selected for a number of reasons:

1. There is a need for such a system. At present about 40% of water delivered by the smallest 1000 public water supplies in the UK falls short of the standards of the EEC directive which took effect in 1985, and these supplies are also about 15 times more costly than the national average[1]. The economic upgrading of these supplies requires the availability of expertise in a useable form, such as can be provided in the form of an expert system.
2. The task was well-defined and did not involve vague or complex reasoning, so that the resulting rule base could be made simple and easy to comprehend.
3. A water engineer, J. Warden from Water Research Centre, was available to serve as the source of expertise for the system. He had already studied the problem of upgrading small water supplies in a methodical manner, and his interest in the characteristics of the general problem aided the knowledge acquisition process greatly.[2]
4. The development project seemed to be of a manageable size. The total development time extended over two years, including construction of the present rule base, and a certain amount of testing by the WRC. The finished product, running on a personal computer, is due to be released to the water industry in April 1987.

The development process

In the absence of any accepted methodology for the development of expert systems, the approach adopted on this project was influenced mainly by a desire to ensure that the user interface to the system should be right. For this reason, it was decided to adopt, as far as possible, a structured methodology for software production based on the data-structure oriented approach of Chen *et al.*[3] In practice, this meant that after an initial and thorough knowledge acquisition phase, in which the logical structure of the system was decided, the project proceeded through the usual phases of a normal software development (see Fig 1). The logical design of the software ensured sufficient flexibility to enable some easy modification of the user interface after the user testing phase. In particular, the modes of user interaction with the knowledge base were improved considerably after testing, by the addition of sensitivity analysis and cost breakdown analysis modules.

It was decided not to build a prototype system using Prolog or Lisp, for example, for the following reasons:

(a) An evaluation as to the practical usefulness of such a system depended greatly upon the quality of the user interface. Although a prototype system could have been built faster using an AI language, such a prototype could not test out crucial features of user acceptability connected with the user interface. Particularly, it was felt that the speed of operation of such a prototype on a personal computer would fatally prejudice any real user acceptability test.

(b) An important consideration in selecting upgrading schemes is one of cost, and the expert task involves a considerable amount of numerical calculation. AI languages are not ideal for such problems, particularly regarding the aspect of speed.

(c) In some respects, prototyping an application using an AI language exactly mirrors the paper logical data design exercise of entity analysis. If the knowledge is not too 'deep' in nature, there is nothing extra to be gained from such a prototype.

Knowledge acquisition

The knowledge acquisition was a collaborative process between the expert in the task from WRC and two knowledge engineers from Thames Polytechnic. The process is illustrated in Fig 2 which shows a cycle during which the expert learnt about expert systems and the knowledge engineers learnt about the task. The common objective was to achieve a formal structure to the task which could be modelled by a suitable inference engine. Six cycles were necessary before an acceptable structure was developed by the knowledge engineers. It is not really certain to what extent this structure represents the expert task before the knowledge engineering exercise and to

---

Thames Polytechnic, London SE18 6PF, UK
(Received December 1986)

0263-0257/87/01027-04$03.00
© 1987 Butterworth & Co. (Publishers) Ltd

no 1
201 3594 3

Knowledge aquisition	Logical software design	Coding	User testing	Refining
6 months	4 months	4 months	6 months	4 months

Fig 1 System development schedule

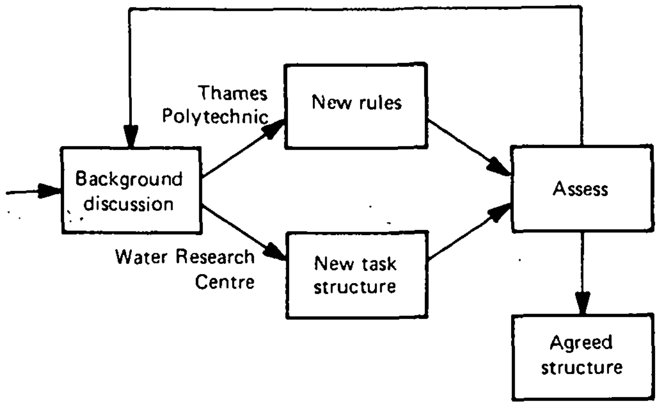

Fig 2 Knowledge acquisition

what extent it was invented from necessity. It does, however, represent the way a methodical expert might work.

In assessing a particular site for upgrading, there are a number of factors to be taken into account. There are analyses of water quality in the form of laboratory reports, environmental factors such as geographical and geological features, local factors, e.g. pertaining to the water authority, and factors relating to the existing supply, e.g. whether suitable buildings and equipment already exist. There are many ways to improve treatment on an existing site, and the expert must use his experience and knowledge to select schemes which will upgrade quality within acceptable cost. First to be described is the type of knowledge brought to bear, and how it is used.

Foremost, the expert will know about the technical aspects of water treatment methods. He will know about filtration and disinfection methods, and about their suitability for different water problems. He knows about which equipment is suitable for sites with only foot access, and which requires vehicular access, as well as which equipment requires electricity on site and which can rely on gravity feed.

But more than just this technical expertise is needed to solve the problem adequately. The expert must use knowledge about the regulations and EEC standards that apply, and about locally sensitive factors such as policy on access road construction or on the use of certain equipment. For the purposes of the computer system, it was also necessary to include some rather obvious common sense knowledge as part of the expert task. For example, it is common sense not to consider expensive schemes when it is known that an inexpensive one will apply; or for instance not to use a filter at all unless it is needed. These common sense rules also have to be stated explicitly for the expert system.

The rule base

In the course of knowledge acquisition as described it became apparent that for this particular expert, the process of selection was one of elimination from a finite set of possible treatments. Although the alternative strategy, i.e. of constructing a set of suitable process units to suit the conditions, was considered, it was thought to be important that all possible treatment methods were to be examined systematically, in case an otherwise optimal solution should be overlooked. This overall selection philosophy gave rise to the architecture of the inference which is illustrated in Fig 3, in which a total set of possible treatment methods is progressively reduced by the application of four classes of rules. This was the structure for the task resulting from the knowledge acquisition process. The basis of this structure is that there is a natural hierarchy to the rules in terms of their relative power in eliminating methods. This relates to the type of knowledge they represent, as follows:

Type 1 rules

These are rules which eliminate some schemes on the grounds of technical feasibility. They state conditions which must be satisfied for a scheme to be considered at all. Usually, they derive from the principle that only certain process flowsheets will solve certain major problems. A group of upgrading schemes can be eliminated if it contains illegal flowsheet combinations. Examples of these rules are:

IF turbidity > 4 NTU THEN use a filter
IF turbidity > 2 NTU AND there is no filtration THEN do not use ultraviolet methods.

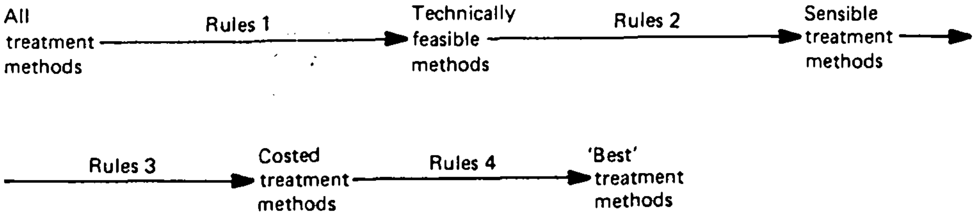

Fig 3 Treatment selection architecture

Type 2 rules

These operate both on data supplied by the user, and on the state of the current set of schemes under consideration. They eliminate schemes on the grounds that it doesn't make sense to carry two schemes through to the next stage if one is certain to be more expensive than the other, all other things being equal. Examples of these rules are:

> IF it is technically feasible to use no filter THEN do not use a filter
> IF electricity does not exist on site AND a constant rate doser scheme is technically feasible THEN do not use ultra-violet methods

These are broad statements about the cost of schemes found technically feasible at stage 1, any of which may be implemented given enough money. It would be possible to pass all the feasible schemes on to costing stage 3, but this would mean a great deal of extra work in evaluating obviously costly schemes.

Type 3 rules

Each of the treatment methods is assumed to have a precise cost C which the expert will know to lie within a broad range $C1 > C > C2$ before he has ascertained any information at all. If he were to perform detailed costing of a scheme, acquiring a maximum amount of information, the cost would be known to be within a finer range $D1 > C > D2$. However, there is an expert phase before the detailed costing which helps to decide the schemes worth the effort involved in detailed costing. This phase refines the broad cost range $C1 > C > C2$ by knowledge of the relationship between key environmental information and the main cost factors.

To model this phase, the system picks upon an important factor f relating to the set of schemes at stage 3. It will then apply the rules which refine the range of values of f. These refer to data supplied by the user, and to the system cost data base. All of this data is uncertain in that it is held or supplied by the user as an upper and lower limit.

Type 4 rules

These represent a final selection phase. In the event that there are a number of similarly priced schemes which will solve a given problem, it is usual for a selection to be made from them on grounds other than just cost. It may be preferrable for instance to select schemes differing as much as possible in concept, and to put these out to tender, rather than select similar schemes. This final phase was not implemented in this application because, on practical grounds, there are usually so few schemes left after the first 3 stages that the final selection can easily be made manually.

The cost model for each scheme takes the general form of a tree connecting all the factors contributing to the total cost, with arithmetical operators forming the nodes of the tree. If the tree were fixed at the outset of this phase, the system would reduce to simple calculation, with the user supplying best estimates for the factors at the leaves of the tree and the system propagating this information up to the root node. Even so, there would be advantages to be gained by expressing the tree as a set of (arithmetical) rules; in fact, the usual advantages of expert systems, e.g. intelligibility, ease of modification and modes of querying the rule base. The tree, however, is not fixed at the outset. There are still many uncertainties concerning the particular needs of the site under consideration regarding necessary equipment, existing buildings etc., which contribute to cost. These uncertainties are built into the system in the form of rules with associated actions to prune off sections from a total cost tree according to information supplied by the user.

As an example, Fig 4 shows a section of the costing tree for the scheme S.CR.V.E. (a scheme using a Slows and filter, a Constant Rate chlorination doser, with Vehicular access and Electrical equipment). The rules represented here are:

(1) S.CR.V.E = reservoir cost + electricity cost + equipment cost + attendance cost + access cost
(2) IF required capacity > present capacity THEN new tank, cost ELSE existing tank cost
(3) new tank cost = tank cost + installation cost
(4) tank cost = required capacity x tank cost factor

Rule (2) establishes whether one needs to evaluate the cost of a new tank or the cost of upgrading the existing tank. It decides by first requesting from the user any data it needs to calculate present capacity and required capacity. According to the result, it prunes off the unwanted branch from the generic costing tree.

The generic tree may be extended to include as much detail as is suitable for the application. As a general principle, it is the level of uncertainty in the data to be provided by the user which determines the degree of detail worth putting into the rule base. The default estimates of cost data also provide a useful database of cost information. Indeed, one view of the system is that it forms an intelligent mechanism for querying this cost database.

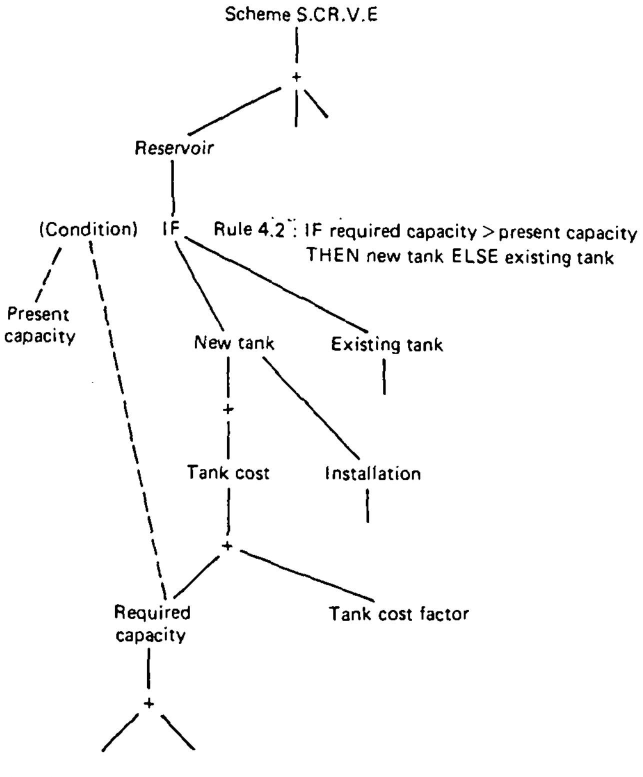

Fig 4 A section of the costing tree

System operation

System modes

In the normal mode of operation, the user is asked to supply facts if necessary during stages 1 and 2, and from a totallity of about 80 possible schemes, not more than about 10 are selected as worthy of further examination. These schemes are displayed at the outset to stage 3, together with upper and lower default costs. e.g.:

Scheme	Lower	Upper
S.CR.V.E	30300	120200
GS.FB.V.E	38800	150300

The user is able to select a scheme for more detailed examination, and is asked to supply both facts, e.g.:

Does electricity exist on site?

and also to improve upon default cost data, e.g.:

Default distance to electricity supply is: Lower 0m, Upper 500m
Do you want to change these?

As the dialogue advances and the system acquires more site information, the upper and lower estimates for all the schemes narrow considerably. At the end of a complete dialogue, the results are available as a printed report on each scheme. The cost is broken down into levels, e.g.:

Cost Breakdown Report

	Lower	Upper
Scheme S.CR.V.E	34500	39200
Reservoir	5400	5900
tank cost	3100	3400
installation cost	2300	2500
Infrastructure	10200	12100
⋮	⋮	⋮

Query modes

There are several ways in which the user is able to query the system. As usual in expert systems, the questions 'why?' at any stage in the dialogue enables the user to look into the rule base at the rule under consideration. And the command 'help' gives access to extra pages of explanatory information concerning the factor currently under examination. This latter facility was used extensively in this application to provide detailed information on the engineering aspects of the total problem.

In addition, it is possible for the user to ask the system why a particular scheme, thought otherwise to be acceptable, is not included in the selected set. This is the query 'why not?', to which the system responds with the particular rule which eliminated the scheme from consideration.

A useful query mode which was added after user testing, is the capability to perform a sensitivity analysis. This allows the user to vary any cost factor and obtain corresponding costings for any selected scheme. The user is thus able to use the system as a modelling tool, asking 'what if?' questions to aid the final choice of the scheme.

Modification mode

The rule base construction and later modification is performed using a separate text editor to produce text files. These contain the rules, default cost data and reporting levels required in the output. The rules are compiled into an internal form to ensure the speed of processing necessary particularly for the number crunching calculations involved in stage 3.

Conclusions

The system, with a fairly complete rule base, has undergone testing both at WRC and independently during demonstrations to water engineers throughout the UK. The results are in line with those produced by human experts and are produced with less effort and time. Typically, a dialogue lasting under an hour can produce an analysis which might otherwise take several days. The system also provides a convenient method for storing knowledge resulting from a research effort into the problems of small water supplies by WRC. It represents both a resource of the organization, and a convenient communication tool for the dissemination of current standards and practice within the water industry.

Other application areas of process selection have the same general structure as that described here, and some consideration is being given to similar problems within the water industry, e.g. for the design of small sludge plants. Other facilities are to be added to the existing shell to enable it to deal with a wider variety of applications. In particular, the facility to handle tables of costs in a simple rule format will enlarge the number of suitable applications considerably.

References

1 **Warden, J.** Small water supplies. *J. Inst. Water Eng. Sci., 1983, 37(4), 313*

2 **Warden, J.** Problems of small water supplies, *J. Assoc. Water Officers, 1986*, **22***(1)*

3 **Chen, P.** The entity-relationship approach to logical data base design, *QED Monogr. Ser. Database Manage. 1977, (6)*