DEPUR: a knowledge-based tool for wastewater treatment plants

[1]Pau Serra, [2]Miquel Sànchez, [1]Javier Lafuente, [2]Ulises Cortés, [1]Manel Poch

[1]Universitat Autònoma de Barcelona
Unitat d'Enginyeria Bioquímica (CSIC – UAB)
Edifici C. E-08193 Bellaterra (Barcelona). Catalonia.
iqte2@cc.uab.es

[2]Universitat Politècnica de Catalunya
Departament de Llenguatges i Sistemes Informàtics (IA)
C/ Pau Gargallo 5. E-08028 Barcelona. Catalonia.
sanchez@lsi.upc.es, ia@lsi.upc.es

Abstract

A malfunction of a wastewater treatment plant is a big social and biological problem. A poorly treated waste water outside the plant could provoke dangerous consequences for human beings as well as the nature itself. The conventional control systems, that are usually applied in this field, have to cope with some difficulties: complexity of the system, an ill-structured domain, qualitative information, uncertainty or approximate knowledge, real-time dynamic system, ... This paper shows an application of artificial intelligence in order to help the operators of wastewater treatment plants in their task of process control. The main goal is to build a knowledge-based tool useful for the diagnosis and management of wastewater treatment plants. First, is made a survey on wastewater treatment plants to describe the complexity of the system being modelled and to outline its own difficulties. It is discussed the development of the application and the methodology employed in it. A new methodology called LINNEO^{+} is introduced. It is used for the automatic knowledge acquisition process in order to build up a Knowledge Base. The prototype's architecture constructed -DEPUR- is detailed together with some obtained results.

Keywords: Knowledge-based Systems, Process Diagnosis and Management, Wastewater treatment, Knowledge Acquisition tools, Environmental Engineering, Biotechnology.

1. Introduction

The aim of this work relies on the application of new techniques in the fields of automation and process control. The AI tools, as seen below, are very important aids in the automation of process management. A malfunction of a wastewater treatment plant is a big social and biological problem. A poorly treated waste water outside the plant could provoke dangerous consequences for human beings as well as the nature itself.

The setting of an automatic process control over the wastewater treatment plant system, has outlined some difficulties:

- The *complexity of the problem*. There are many facts from different nature present at the system: chemical facts, mechanical, biological, microbiological ones, …
- A *non-well structured domain*. The relationships among the concepts or attributes of the domain are not enough known or the agreement among experts is not very high.
- *Most information is neither numeric nor quantified*. So, it is not suitable to be included in the context of a classical -numerical- control model (microbiological information,…).
- *Uncertainty or approximate knowledge*. By example, the subjective information supplied by the experience of the plants' managers.
- A *dynamic system*. Continuous changes that modify directly the performance of the process.

Moreover, it can be said that the *classical methods* work fine on the *normal* states of the plant but not in other *abnormal* states of working. How could they detect some mechanical faults?. How could they use the available information but incomplete one, to solve a specific problem?. These troubles with the conventional process control systems as *feed-back control*, *feed-forward control*, *optimal control*, *adaptive control*, … have arisen during last years much research effort in Artificial Intelligence [1–4].

Bringing to mind the main advantages and characteristics of knowledge-based systems, specially those of expert systems, and the above difficulties one could think that match very well. So it has been developed a prototype of a *qualitative expert system with uncertainty reasoning* and a *real time numerical control expert system* by the other hand, which is described in [5], for the diagnosis and management of wastewater treatment plants.

2. A wastewater treatment plant

The main goal of a wastewater treatment plant is to reduce the level of pollution of the wastewater at the lowest cost, that is, to remove -within possible measure- strange compounds (pollutants) of the inflow water to the plant prior to discharge to the environment. So, the effluent water has low levels of the pollutants (lower than maximum

ones allowed by the law). The plants taken as models are based on the main biological technology usually applied: *the activated sludge process* [6]. The concrete wastewater plant studied is located in Manresa, near Barcelona (Catalonia). This plant receives about 35000 m^3/day inflow from 75000 inhabitants.

A wastewater treatment plant is composed usually by two stand-alone but interactive subsystems: *sludge line* and *water line*. The study has focused on the water line. It is formed by a sequence of unitary processes where the effluent of one process becomes the inflow to the next one. Usually there are 3 major processes: primary treatment, secondary treatment and tertiary treatment. Each one reduces or removes the concentration of several specific pollutants. The chart of a plant may be seen in figure 1.

Figure 1 here

3. Development of the qualitative expert system

It has been built up an expert system, called DEPUR [7–8], in order to support the approximate or uncertain reasoning, qualitative knowledge, ... in the managing of wastewater treatment plants. The main problem appearing on a wastewater treatment plant is the breakdown of the balance between *biomass* and *substrate*, as shown in figure 2. The biomass is formed by a set of microbiological beings that in presence of oxygen, degrade non-sedimentable organic suspended solids (i. e., the substrate) in the water. More information about the wastewater treatment processes can be found in classical bibliography [9–10].Thus, the task of diagnosis of the system tries to determine the causes of the problems detected. Often, the most serious problems are related to the biological process in the secondary treatment.

Figure 2 here

The design process of a knowledge-based system can be expressed commonly as shown in figure 3. The work has been focused on the water line subsystem domain. The *relevant attribute selection* process has been developed twofold: firstly, in a typical manner based on the attribute selection made by the experts of the domain, and, secondly, by means of a new automatic attribute selection approach based on the relevance and potential utility concepts [11].

Figure 3 here

The *attribute structuring* process has been made over two approaches:

- *Decision trees*: the classical approach derived from the interactions between the expert and the engineer. This approach is governed by the cause-effect criterion. Multi-valued decision trees have been used and inference rules are

obtained from them. The rules point out to discover the causes of a certain problem or situation previously detected.

- *Automatic classification techniques*: used as an automatic knowledge acquisition tool is an approach recently derived from machine learning in Artificial Intelligence. The LINNEO+ system, based on fuzzy non-well structured domain classifications, has been used in this phase. From the characterization of the obtained classes, more general inference rules are constructed. They lead to know the most likely working situation of the plant. This methodology will be explained more accurately in next section.

Decision trees have been used in the causes identification phase and automatic classification techniques in the problem identification process as shown in figure 4.

Figure 4 here

4. The methodology LINNEO+ for the knowledge acquisition process

The software LINNEO+ [12–13] is a knowledge acquisition tool useful in ill-structured domains which uses both *inductive learning methods* and *deductive* ones. LINNEO+ uses -possible noisy- observations to induce classes as well as taking advantage of the expert's current state of knowledge (which may be incomplete), expressed as logical constraints on the possible output of the classification process.

Figure 5 shows the general framework of the methodology. The human expert abstracts a collection of observations (sample) from the problem domain that he thinks of as relevant. He also selects a set of attributes (descriptors) that he assumes to be *significant* for the characterization of the domain objects. A sample is then modelled by a set of attribute/value vectors. At the same time, the expert can express what he already knows about the domain. This previous knowledge about the domain (that will be called as Domain Constraints -DC-) is represented as constraints and examples on classes already known to exist.

Figure 5 here

Starting with this knowledge and information, LINNEO+ produces a classification. It uses induction over the observations taking into account the DC which guide the inductive process. The results of this classification are an intensive and extensive description of the generated classes and a fuzzy membership matrix that relates observations to generated classes. The expert evaluates the results and possibly modifies some attributes or some DC.

When a "correct" classification is obtained, rules are derived from the generated classes. Each rule includes in its antecedent the membership conditions and a fuzzy certainty factor

for a particular class. Each antecedent is a conjunction of logical conditions on the values that descriptors must show for belonging to the class whose identifier appears on the rule's consequent. These rules are analysed and, through subsumption detection and synonymy analysis, they are organized in a hierarchical way. The results are made known to the user, who can accept and confirm them. He also has the chance to go back to the classification module, to the definition of the DC or to the sample selection step and obtain other results. When all these actions are over, rules can be validated using new observations not previously in the sample. As soon as the expert accepts the resulting rule set, it is transformed into IF-THEN rules and modules.

5. Architecture of the application

The modelling of the domain has been made as follows: the *objects or observations* are prototypical working situations or problems in the wastewater plant of Manresa, near Barcelona (Catalonia). The experts have defined about 19 different situations in the beginning of the work. The *attributes or properties* are the most relevant parameters and facts -in the viewpoint of the experts- that characterize the domain. They have selected about twenty attributes. So, the input matrix to the classification process of LINNEO$^+$ is as follows:

• Objects -situations- defined:

- Normal
- Bulking-non-filamentous
- Bulking-not-enough-oxygen
- Bulking-sulphurs
- Rising
- Foaming
- Overaeration
- Summer
- Winter
- Storm
- High plant inflow
- Rotation-band-crash
- Primary-sludge-sedim.-problem
- Not-enough-primary-sludge-exit
- Overloading
- Turbine-crash
- By-pass-bad-closed
- Toxic-substances-loading
- Toxic-substances-increase

• Attributes defined:

- Rotate-band-clarifiers
- Bubbles-clarifiers
- Bubbles-primary-settlers
- Inflow-water
- Dissolved-Oxygen
- Inflow-DQO
- Outflow-DQO
- Sludge-age
- Foam-bassins
- Sulphurs
- Sludge-exit-primary-settler
- Suspended Solids-bassins
- Suspended Solids-inflow
- Suspended Solids-outflow
- Volatile Suspended Solids-bassins
- Volatile Suspended Solids-recirculat.
- Temperature
- Turbines

- Filamentous
- SVI
- Odor-bassins
- Toxic-substances-bassins
- Toxic-substances-inflow

An example of a complete observation with its values for the attributes may be like this one:

• Bulking-non-filamentous:

Outflow-DQO	——>	High
Sludge age	——>	Old
Filamentous	——>	Normal
SVI	——>	High
Volatile SS-recirculation	——>	Low
All-other-attributes	——>	Nought-value

The other attributes have *nought* values [14] or *don't care* values. It means that the values of these attributes are not relevant for the characterization of the object. Although they could be relevant to describe other working situations of the plant. For instance, in the above example, the value of the attribute *temperature* could be anyone -but don't care *which* is- for the description of the bulking situation originated by non-filamentous causes.

5.1 Results

10 classes have been obtained (see figure 6) as a result of the classification process. Each one, groups closely related prototypical situations. The classification shows the most common situations and problems in the working state of the wastewater treatment plant and makes easier the arrangement of the underlying domain. These classes lead the inference process in the diagnosis.

Figure 6 here

6. Implementation of DEPUR in MILORD

MILORD [15–16] is a classical second generation expert system shell, but considerations have been made on some important features related to this particular domain, as the microbiological and temporal ones. The application developed is called DEPUR and reflects the experts knowledge. See figure 7. The *Data Base* is a representation of the current state of the system, that could include some past or future facts of interest. In this representation are selected the attributes more relevant for the domain characterization. The *dictionary* of the Data Base is composed by 167 attributes with a decrease by means of relevant and potential utility concepts. An example of dictionary entries may be:

(INFLOW-WATER-NORMAL
"Has the inflow water normal values (35000 – 45000) m^3/day ?"

```
 FUZZY)

(LOW-TEMPERATURE
 "Is the temperature low (< 5) ºC ?"
 FUZZY)
```

The *Knowledge Base* is the arrangement of the knowledge you have over the domain, made explicit by the expert and by information coming from the analysis of the system being modelled. Usually, this knowledge is expressed by means of *inference rules*. Examples of rules may be:

```
(RMENU003B
      (IF    (DQO-EFLUENT-WATER-HIGH
              SLUDGE-AGE-YOUNG))
             SURE
      (THEN
             (INFER     (CLASS-3B))))

(RC3004
      (IF    (CLASS-3A
              DISSOLVED-OXYGEN-LOW
              ODOR-BASSINS))
             SURE
      (THEN
             (INFER     (BULKING-O2))))
```

The system is formed by 79 rules, grouped in 12 modules. A module is a structural element of the Knowledge Base, composed by a set of related rules.

Figure 7 here

The *Control Module* has the goal of control *how* and *when* may be used the knowledge supported by the system. Often it is expressed by means of metarules. The prototype has got 15 metarules. An example of a *metarule* may be:

```
(M003B
      (IF    (CLASS-3B))
             SURE
      (THEN
             (TREAT-MODULES  C3)))
```

The linguistic terms used to manage the approximate knowledge are in order of certainty:

Sure
Almost-sure
Very-possible
Possible
Moderately-possible
Slightly-possible
Almost-impossible
Impossible

The modularity of the application, the self-explanation and the user friendly interface afforded by the shell are other outstanding properties.

7. Validation

The prototype diagnostics in contrast to the answers of human experts, has yielded very good results. It has been tested with several prototypical situations in the working of the plant (bulking, foaming, …).

Currently, this classification is validated against a data stream daily obtained from direct observation in the wastewater treatment plant during the last 2 years, with very high degree of performance.

A short dialogue with the application could be like the following one:

> *Has the effluent DBO high values (25 – 40) mg/l ?*
– Sure
> *Has the water inflow high values (45 000 – 55 000) m^3/day ?*
– Almost impossible
> *Is the temperature low (< 5) ºC ?*
– Impossible
>*Has the inflow water high concentration of Zn (4 – 20) mg/l ?*
– Impossible
> *Is the Volumetric index of sludge (SVI) high (150 – 300) mg/l ?*
– Very possible
> *Are the Suspended Solids (SS) in the recirculation low (1300 – 4600) mg/l ?*
– Sure
> *Are the Volatile Suspended Solids (SSV) in the recirculation low (1000 – 3500) mg/l ?*
– Sure
> *Has the dissolved oxygen (DO) high values (2.5 – 10) mg/l ?*
– Impossible
> *Is the age of sludge old (8– 20) days ?*

– Impossible

. . .

> Conclusion —> Bulking-not-enough-oxygen — Very-possible

. . .

8. Conclusions and future work

The building of DEPUR system as a prototype of fuzzy expert system applied to managing a wastewater treatment plant, has been the consequence of the research effort during last years (study and modelling of wastewater treatment plants, development of new techniques of selection and arrangement of attributes, building of computer systems, …) by both scientific groups involved.

A diagnosis and managing tool based on knowledge-based systems for a wastewater treatment plant may be a useful approach in the field of intelligent systems applied to managing industrial processes.

The prototype may be helpful for the manager of the plant. The treatment of uncertainty and the new fuzzy techniques for automatic classification are some of its characteristics. This methodology has been successfully applied to other domains. See [11–12].

This project supposes a step towards the building of intelligent systems oriented to preserve the water, as part of the nature (ecological intelligent systems).

Other interesting point is the possibility to adapt the system to the specific characteristics of the concrete wastewater treatment plant under control.

In the future, it is proposed to reach a second step: the development of a Supervisor Knowledge-Based System -more powerful than an expert system-, that join diagnosis and control including the numerical control algorithms [17] with qualitative and heuristic information and reasoning. See figure 8.

Also there is much work to do: extend the domain to the sludge line, research in a more accurate way the microbiological compound of the biological process and make a better temporal information treatment.

Figure 8 here

Acknowledgements

This work has been supported by the "Junta de Sanejament de la Generalitat de Catalunya" and the Spanish CICyT projects ROB89-0479-C03-02, ROB91-1139 and TIC91-1087-C03-01. Also the authors wish to acknowledge the cooperation of Ricard Tomàs, manager of the Manresa's wastewater treatment plant.

References

1. J.J. Coleman, S. Krovvidy, R. Scott Summers, W.G. Wee. "An AI Approach for Wastewater treatment systems". *Journal of Applied Intelligence* vol. I, num. 3, pp. 247-261. Kluwer Academic Publishers (1991).

2. V. Floris. "Artificial Intelligence in the operation and management of water resources in South Florida". Personal Communication (1989).

3. K. Maeda. "A knowledge-based system for the wastewater treatment plant". *Future Generation Computer Systems* 5, pp. 29-32. North Holland (1989).

4. R. Sanz, J, Aguilar, C. Sierra, Ll. Godó, A. Ollero. "Adaptive control with a supervisor level using a rule-based inference system with approximate reasoning". *Artificial Intelligence in Scientific Computation: towards 2nd generation systems*. IMACS (1989).

5. P. Serra , M. Sànchez, J. Lafuente, U. Cortés, M. Poch. "ISCWAP: a knowledge based system for supervising activated sludge processes". Submitted (1993).

6. J. Robusté. "Modelització i identificació del procés de fangs activats". Ph. D. Thesis. Departament de Química. Universitat Autònoma de Barcelona (1990).

7. M. Sànchez, P. Serra, Ll. Belanche, U. Cortés. "A knowledge-based system for the diagnosis of waste-water treatment plants". *Proc. of 5th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-92)*. Springer-Verlag. *Lecture Notes in Artificial Intelligence* 604, pp. 324-336. Paderborn, Germany. (1992).

8. M. Sànchez. "DEPUR: Aplicació dels sistemes basats en el coneixement al diagnòstic en plantes de tractament d'aigües residuals". Master Thesis. Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya (1991).

9. L.D. Benefield, C.W. Randall. *Biological process design for wastewater treatment*. Prentice-Hall. N.J. (1980).

10. Metcalf & Eddy Inc. *Wastewater engineering: treatment/disposal/reuse*. Mc Graw-Hill, 3th edition (1991).

11. Ll. Belanche, U. Cortés. "The nought attributes in knowledge-based systems". *Proc. of the European workshop on Verification & Validation, EUROVAV-91*, pp. 77-103. Jesus College. Cambridge. UK. (1991).

12. J. Béjar, U. Cortés. "LINNEO^{+}: Herramienta para la adquisición de conocimiento y generación de reglas de clasificación en dominios poco estructurados". *Actas 3er Congreso*

Iberoamericano de Inteligencia Artificial, IBERAMIA-92, pp. 471-481. La Habana, Cuba (1992).

13. M. Martín, R. Sangüesa, U. Cortés. "Biasing induction with previous knowledge for knowledge acquisition in imprecise domains". *Proc. Avignon '91*, pp. 359-370. 1991.

14. E. Sánchez. "Importance in knowledge systems". *Information System* 14 (6), pp. 455-464 (1989).

15. L. Godó, R. López de Mántaras, C. Sierra, A. Verdaguer. "MILORD: the architecture and management of linguistically expressed uncertainty". *Int. Journal of Intelligent Systems* 4, pp. 471-501 (1989).

16. C. Sierra. "MILORD: arquitectura multi-nivell per a sistemes experts en classificació". PH. D. Thesis. Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya (1989).

17. R. Moreno, C. de Prada, J. Lafuente , M. Poch, G. Montague. "Non-linear predictive control of dissolved oxygen in the activated sludge process". In *Modelling and Control of biotechnological processes*, pp. 289-293. Ed. M.N. Karim, G. Stephanopoulos (1992).

Fig. 1 - Chart of a wastewater treatment plant

Fig. 2 - Biological balance

Fig. 3 - Design process of a Knowledge-Based System

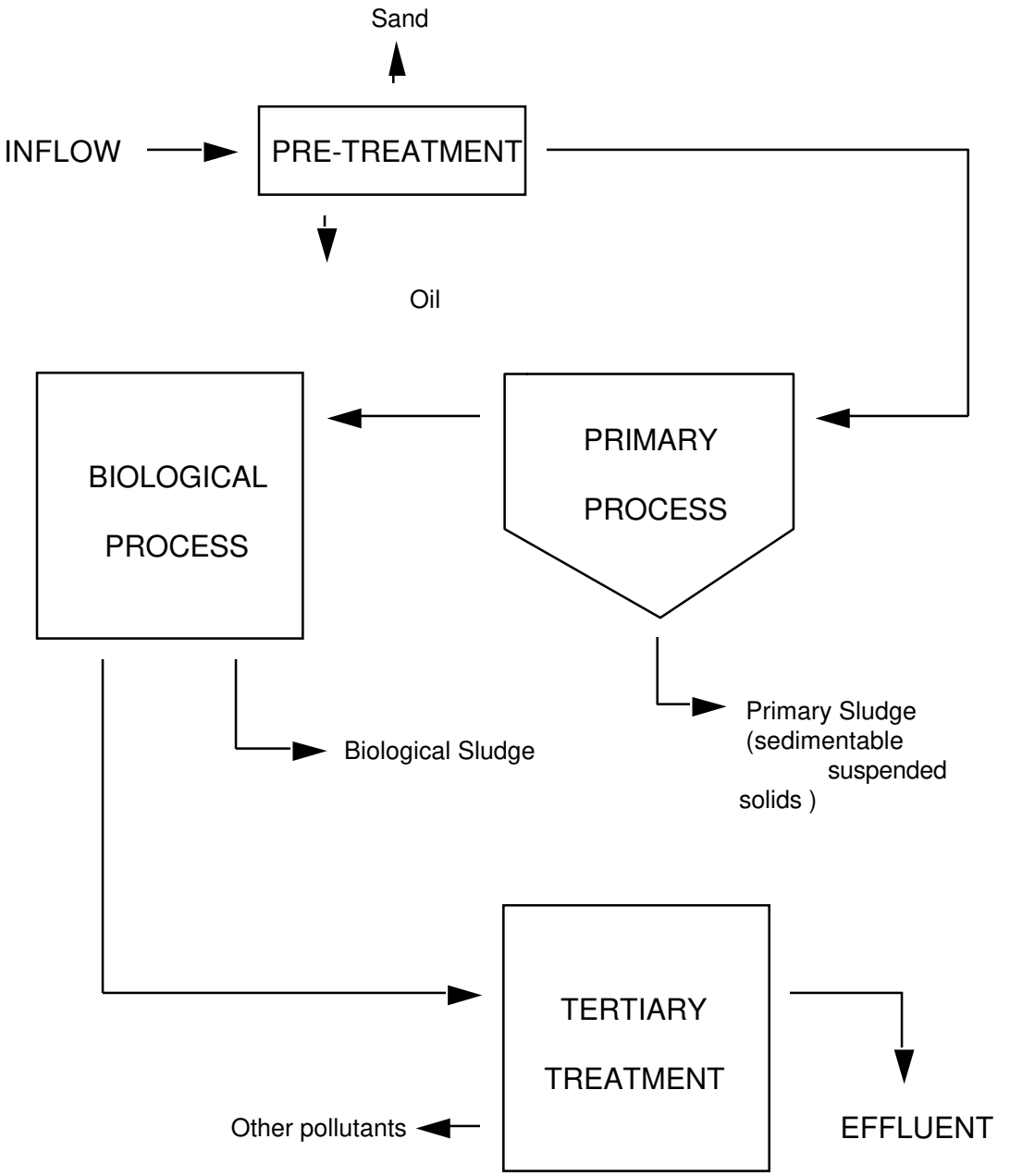
Fig. 4 - Diagnostic process in an Expert System

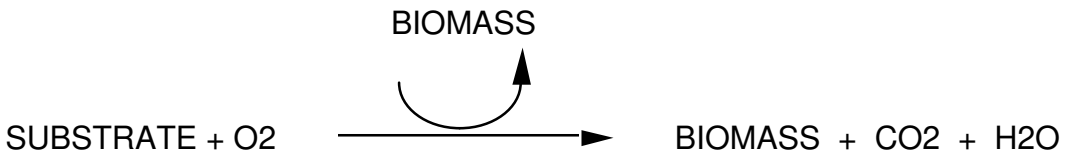
Fig. 5 - The methodology of LINNEO+

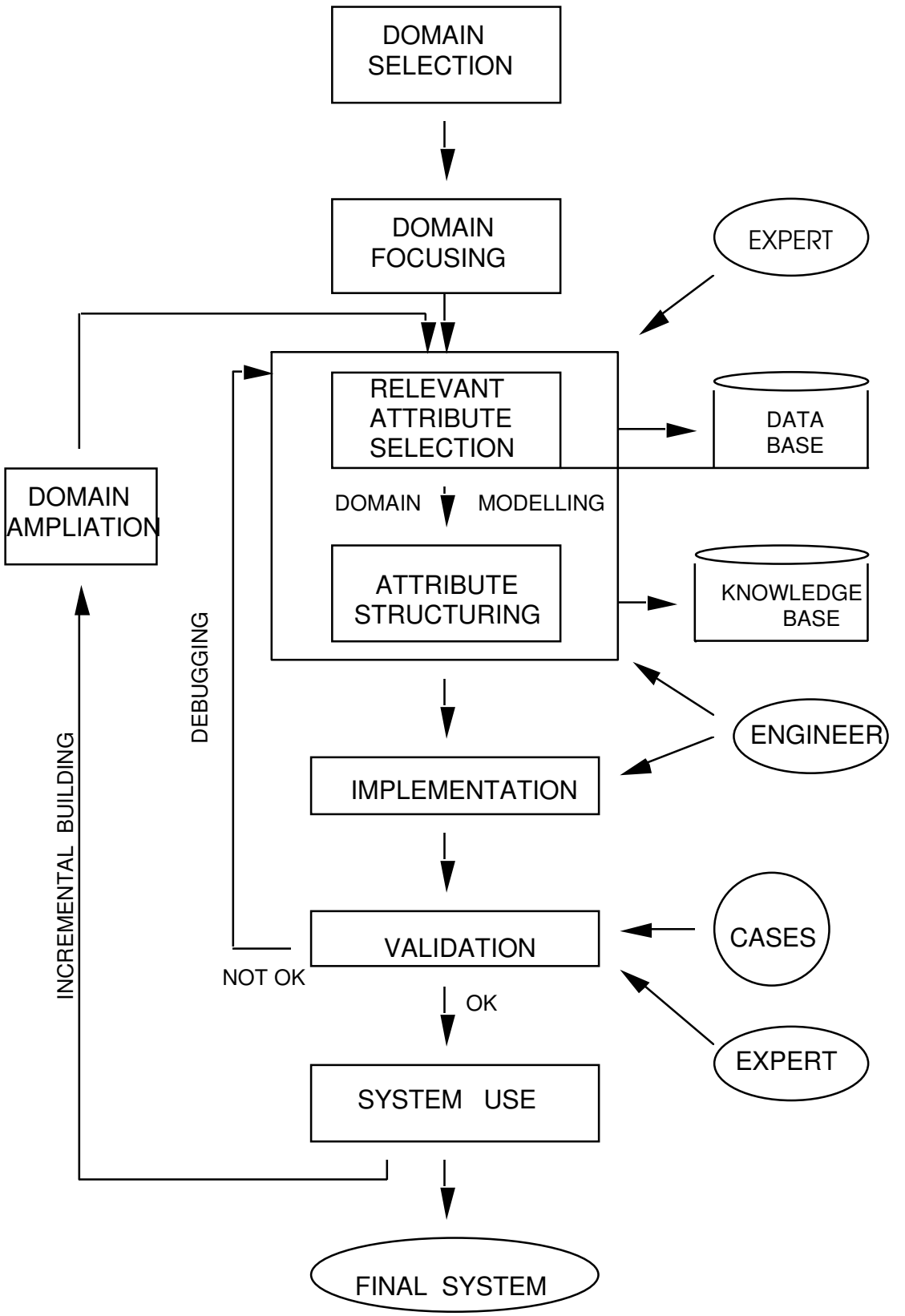
Fig. 6 - Results of the classification process

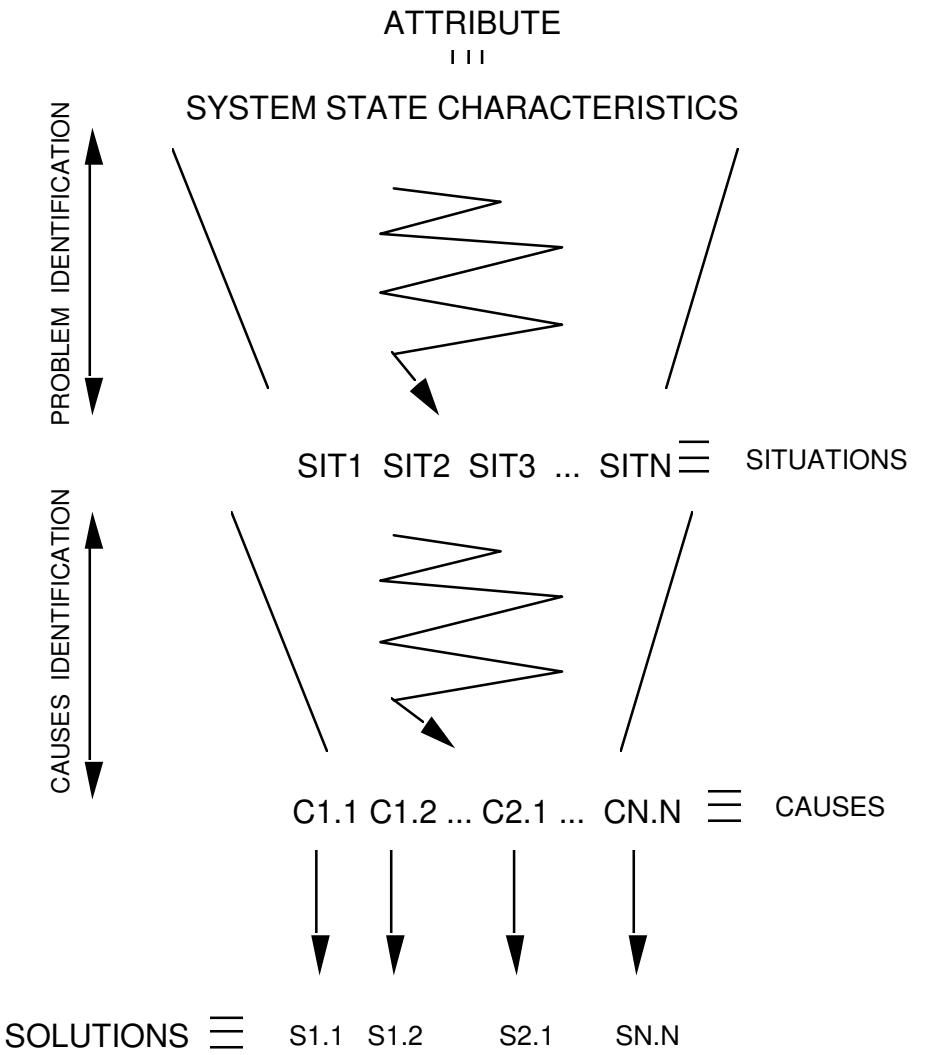
Fig. 7 - General architecture of DEPUR

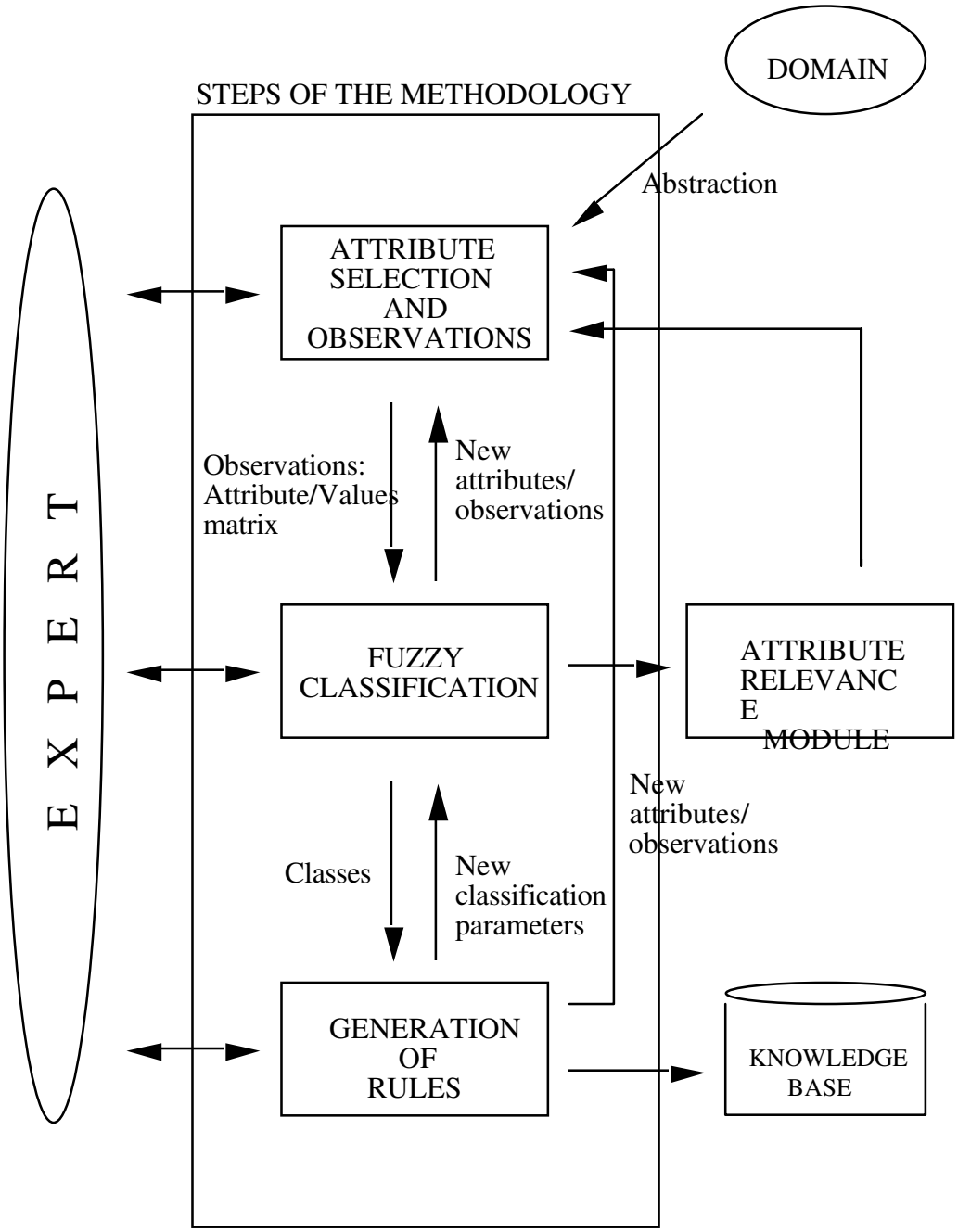
Fig. 8 - A Supervisor Knowledge-Based System

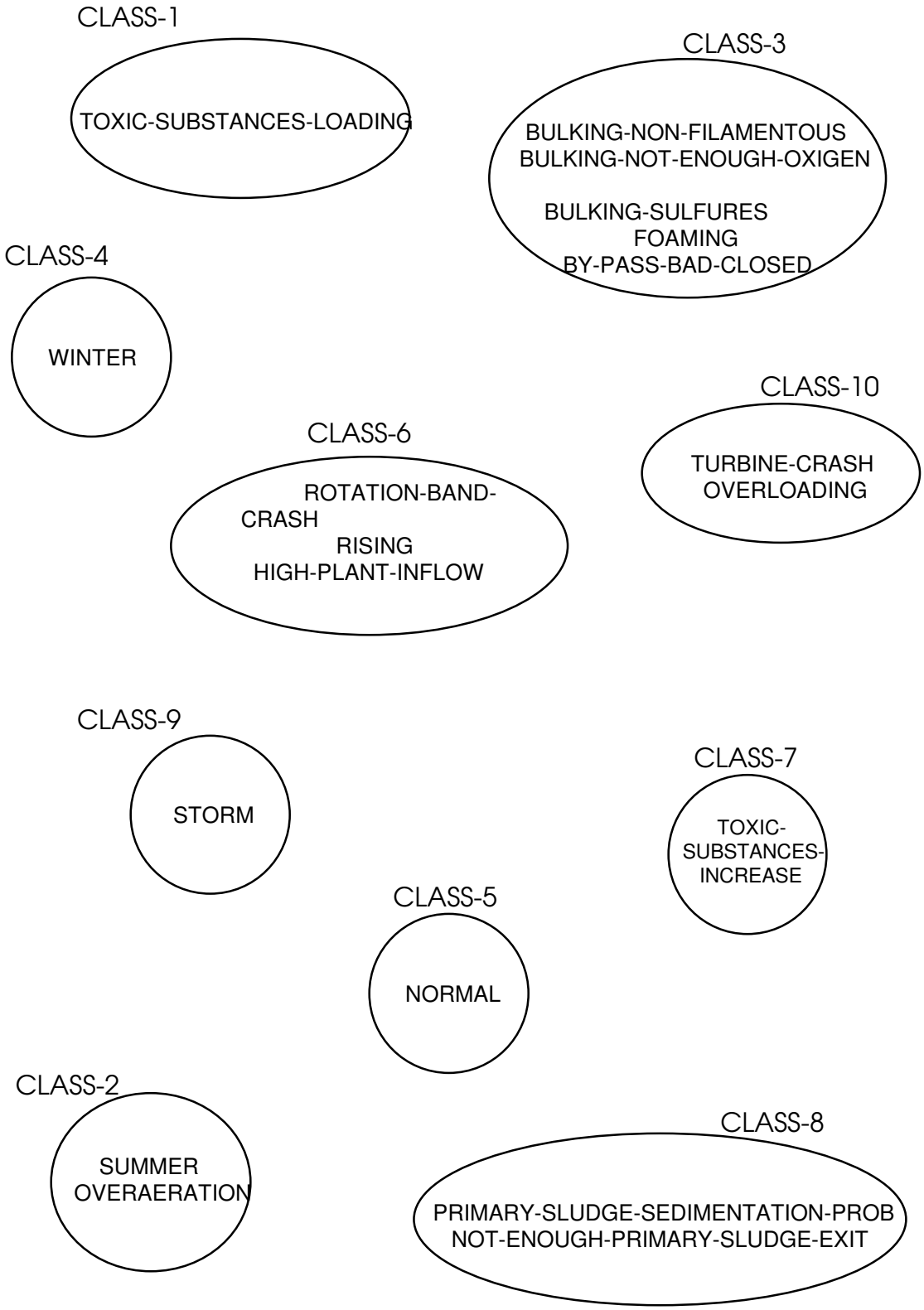

Sand
INFLOW
PRE-TREATMENT
Oil
BIOLOGICAL
PROCESS
PRIMARY
PROCESS
Biological Sludge
Primary Sludge
(sedimentable
suspended
solids)
TERTIARY
TREATMENT
Other pollutants
EFFLUENT

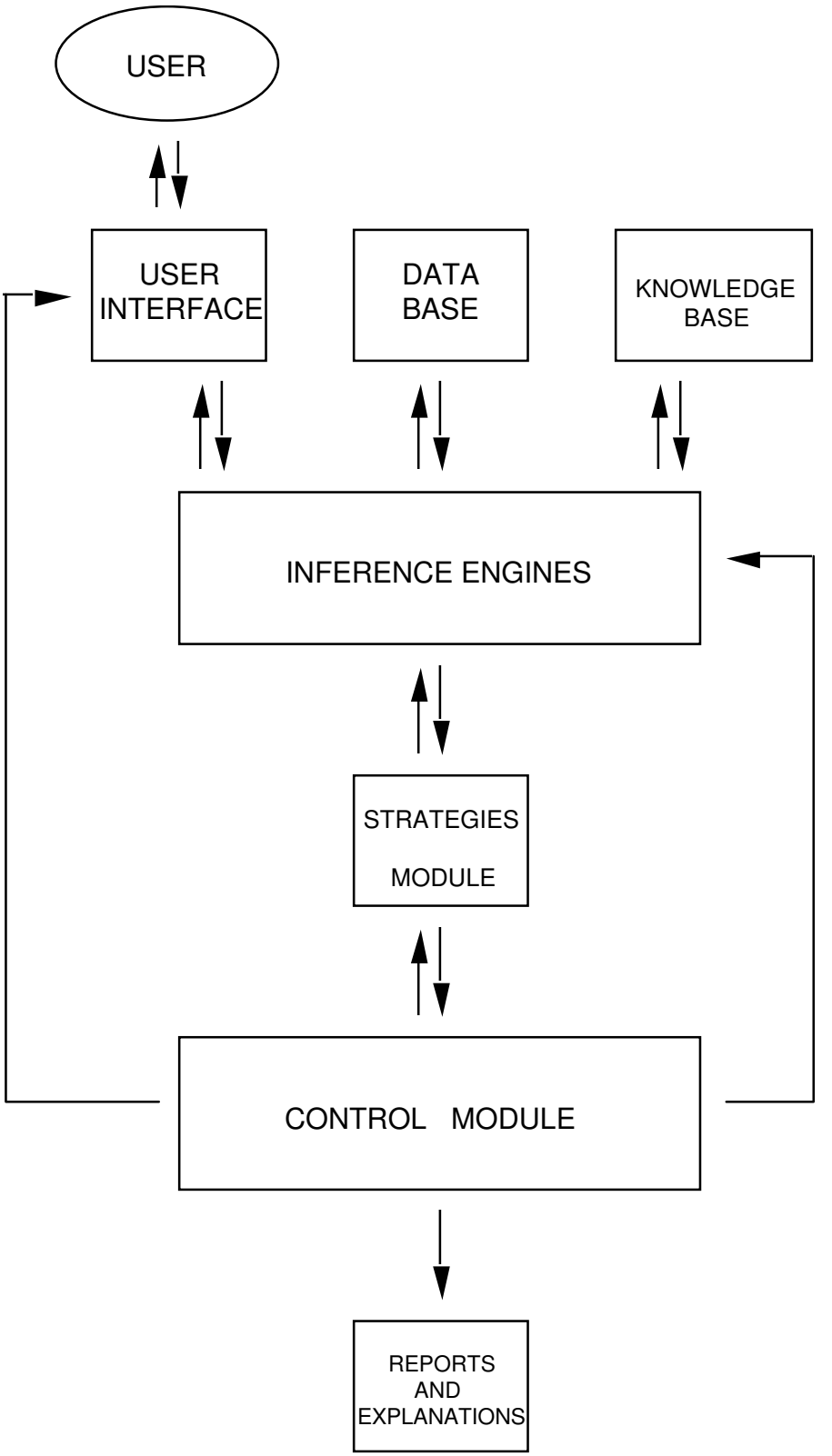

BIOMASS
SUBSTRATE + O2
BIOMASS + CO2 + H2O

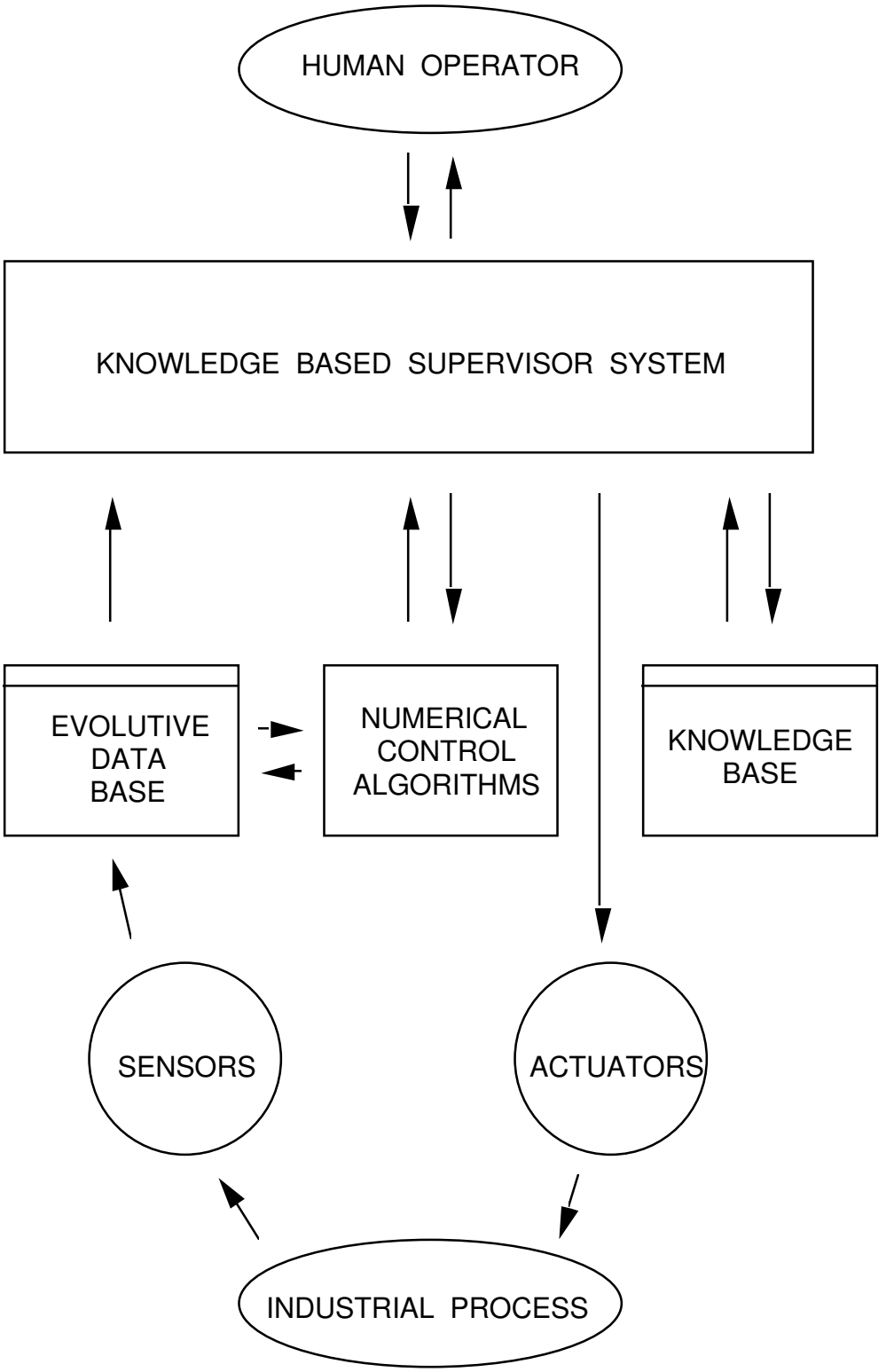

DOMAIN SELECTION
DOMAIN FOCUSING
EXPERT
RELEVANT ATTRIBUTE SELECTION
DATA BASE
DOMAIN AMPLIATION
DOMAIN MODELLING
ATTRIBUTE STRUCTURING
KNOWLEDGE BASE
DEBUGGING
ENGINEER
IMPLEMENTATION
INCREMENTAL BUILDING
VALIDATION
CASES
NOT OK
OK
EXPERT
SYSTEM USE
FINAL SYSTEM


ATTRIBUTE
SYSTEM STATE CHARACTERISTICS
PROBLEM IDENTIFICATION
SIT1 SIT2 SIT3 ... SITN
SITUATIONS
CAUSES IDENTIFICATION
C1.1 C1.2 ... C2.1 ... CN.N
CAUSES
SOLUTIONS
S1.1 S1.2 S2.1 SN.N


DOMAIN
STEPS OF THE METHODOLOGY
Abstraction
ATTRIBUTE SELECTION AND OBSERVATIONS
Observations: Attribute/Values matrix
New attributes/ observations
FUZZY CLASSIFICATION
ATTRIBUTE RELEVANCE MODULE
New attributes/ observations
Classes
New classification parameters
GENERATION OF RULES
KNOWLEDGE BASE
E X P E R T


CLASS-1
TOXIC-SUBSTANCES-LOADING
CLASS-3
BULKING-NON-FILAMENTOUS
BULKING-NOT-ENOUGH-OXIGEN
BULKING-SULFURES
FOAMING
BY-PASS-BAD-CLOSED
CLASS-4
WINTER
CLASS-10
TURBINE-CRASH
OVERLOADING
CLASS-6
ROTATION-BAND-CRASH
RISING
HIGH-PLANT-INFLOW
CLASS-9
STORM
CLASS-7
TOXIC-SUBSTANCES-INCREASE
CLASS-5
NORMAL
CLASS-2
SUMMER
OVERAERATION
CLASS-8
PRIMARY-SLUDGE-SEDIMENTATION-PROB
NOT-ENOUGH-PRIMARY-SLUDGE-EXIT


USER
USER INTERFACE
DATA BASE
KNOWLEDGE BASE
INFERENCE ENGINES
STRATEGIES MODULE
CONTROL MODULE
REPORTS AND EXPLANATIONS


HUMAN OPERATOR
KNOWLEDGE BASED SUPERVISOR SYSTEM
EVOLUTIVE DATA BASE
NUMERICAL CONTROL ALGORITHMS
KNOWLEDGE BASE
SENSORS
ACTUATORS
INDUSTRIAL PROCESS