Integrating General Expert Knowledge and Specific Experimental Knowledge in WWTP

Miquel Sànchez and Ulises Cortés
Dept. de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Carrer Pau Gargallo 5
08028 Barcelona. Catalonia
{sanchez,ia}@lsi.upc.es

Ignasi R.-Roda and Manel Poch
Laboratori d'Enginyeria Química i Ambiental
Universitat de Girona
Plaça Hospital 6
17071 Girona. Catalonia
rqte3@onyar.udg.es

Abstract

The development of an architecture able to manage efficiently the different elements of the process (*integrated architecture*), to learn from past experience (*specific experimental knowledge*) and to acquire the domain knowledge (*general expert knowledge*) are the key problems in real-time control AI systems design. These problems increase when the process belongs to an *ill-structured domain* and it is composed by several complex operational units. Therefore, an integrated AI methodology which combines both kind of knowledge is proposed. This multi-paradigm reasoning provides the target system –a wastewater treatment plant (WWTP)– with some advantages over other approaches applied to real world systems.

1 Introduction

1.1 Wastewater Treatment Plants and AI

The main goal of a wastewater treatment plant is to reduce the pollution level of the wastewater at the lowest cost, that is, to remove –within the possible measure– strange compounds (pollutants) of the inflow water to the plant prior to discharge to the environment. So, the effluent water has the lower levels of pollutants as possible (in any case, lower than the maximum ones allowed by the law).
The plants taken as models –in this study– are based on the main biological technology usually applied: *the activated sludge* process [Robusté, 1990]. The target wastewater plant studied is located in Manresa, near Barcelona (Catalonia). This plant receives about 30000 m^3/day inflow from 75000 inhabitants. The activated sludge process directly depends on live beings (microorganisms), and therefore, on changes experimented by them. It could be possible to get a good plant operation if the supervisory control system is able to react to the changes and deviations of the system and can take the necessary actions to restore the system's performance.

These features reveal that supervision and control of activated sludge processes, could only be treated in a multi-disciplinary integrated way [Venkatasubramanian, 1994], that includes: *monitoring* (sensor developing, continuous analysis equipment), *modelling* (equations that model the bioreactors' behaviour), *control* (maintaining good effluent water quality and reducing operation costs), *qualitative information* (microbiological information, water's colour and odour, water's appearance, etc.), *expert knowledge* (supplied by the large experience from plants' managers, biologists and operators) and *experimental knowledge* (specific knowledge supplied by the previous solved problems in the system). Last features, commonly provide the systems with incomplete, uncertain or approximate information. Therefore, AI can take part a good role in WWTP supervision [Stephanopoulos, 1990; Patry and Chapman, 1989; Stephanopoulos and Stephanopoulos, 1986].

1.2 Goals

This paper presents an approach to integration of general knowledge (coming from experts' domain knowledge or expert knowledge) with specific knowledge supplied from previous solved problems in the system (experimental or practical knowledge). The expert knowledge will be modelled through explicit inference rules, while the experimental knowledge will be modelled by means of cases or experiences. This cooperation tries to get benefit from the advantages of both kinds of knowledge, and to cope with typical shortcomings either from knowledge-based systems: do not learn from experience, the knowledge acquisition problem, the brittleness; or from automatic control systems: complexity of the processes, ill-structured domains, non-numerical or qualitative information, uncertainty or approximate knowledge. This paradigm combination integrates in a single architecture some cognitive processes as knowledge-based reasoning, case-based reasoning, learning, knowledge acquisition, problem solving [Plaza *et al.* , 1993].

2 General expert knowledge

In this paradigm, the knowledge about the domain is modelled with inference rules. The process of extracting these rules (knowledge acquisition) from experts –with interactive sessions where the experts try to make explicit their knowledge and reasoning processes– is quite often described as a bottleneck. Therefore, much effort in AI has been addressed to overcome it. In this case, the knowledge acquisition process have been done with Linneo$^+$ [Béjar, 1995] (a semi-automatic classification tool). The main objective of Linneo$^+$ is to build classifications for ill-structured domains; where much imprecise information exists, it is assumed that observations vary in their degree of membership with regard to each class. Bearing all this in mind, the use of the conventional concept of distance as a fuzzy similarity value is used.

A crucial point for the WWTP supervision is the concept of working situations. *A **situation** is an operational working state of the plant, described by measures of the relevant attributes of the process*. The experts characterized 20 *working situations* of a plant using 23 attributes [Serra *et al.*, 1994]:

bulking-non-filamentous	normal
bulking-sulphures	rising
bulking-not-enough-oxygen	foaming
bulking-too-much-oxygen	underaeration
bulking-F/M	overareation
bulking-toxic-substances	surging
high-plant-inflow	storm
bad-clarifiers-operation	overloading
bad-primary-settlers-operation	bad-wasting
in-plant-overloading	
toxic-substances-loading	

Each situation could be defined in terms of raw descriptions and relationships. For example, the *bulking-non-filamentous* situation was defined as follows:

Outflow-COD	–> High
Sludge age	–> Old
Filamentous	–> Normal
SVI	–> High
Volatile SS–recircul.	–> Low
All-other-attr.	–> Nought-value (don't care)

The results of the classification process with Linneo$^+$ provide intensive and extensive description of the generated classes (*situations*) and a *fuzzy membership matrix* that relates observations (data) to generated classes. At this point, inference rules can be generated. This process is depicted in figure 1. Bearing in mind the *prototype* of a class and the superclassification structure, inference rules that lead the *diagnosis* process in the target system, can be derived. These rules *identify* to which class (set of situations) belongs a given observation (data). *Identifying rules* characterize the values that descriptors of new data must show for being member of a class. For instance a rule as the following could be generated:

```
(if     (COD-EXIT  HIGH
         SVI  HIGH
         SS-RECIRCULATION  LOW
         SSV-RECIRCULATION  LOW)
    VERY-POSSIBLE
           (INFER  CLASS-3A))
```

Other *discriminating rules* must be provided to discriminate to which *situation* –within a class– belongs a given observation. For this task is taken into account the *fuzzy membership matrix*. For example:

```
(if     (CLASS-3A
         SLUDGE  OLD
         FILAMENTOUS  NORMAL)
    ALMOST-SURE
   (INFER  BULKING-NON-FILAMENTOUS))
```

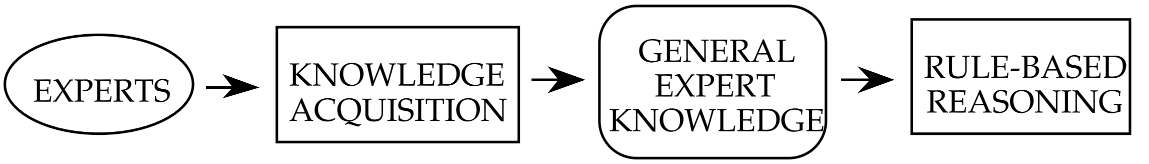

Figure 1. General expert knowledge

This process is semi-automatic, so that requires the experts final validation. All these rules are analyzed (subsumption detection, synonymy analysis, etc.) and, afterwards, they are organized in a hierarchical way to guide the diagnosis process. The results are made known to the experts who can accept and confirm them, or have the chance to go back to the classification process.

When all these actions are over, rules can be validated using new observations not previously included in the classification sample. As soon as rules are accepted by experts, they can be incorporated into the KB's of the distributed architecture. These rules capture the subjective domain knowledge of the experts in their daily work at the WWTP. From the several KB's can be done the diagnosis process that leads to identify the *generic working situation(s)* of the plant.

3 Specific experimental knowledge

In this approach, the knowledge about the practical problem solving in the domain is represented by means of cases or experiences (in this case *situations*), which are organized in the Case library. This Case library contains information about previously detected situations and solutions given to them as well as their efficiency (specific experimental knowledge). A Case-based reasoning can be performed in order to get benefit from these past experiences and cases. The Case library will be modified accordingly with the new information [Kolodner, 1993; Schank and Slade, 1991; Riesbeck and Schank, 1989].

Cases denote a *working situation* of the plant. These cases are previously experienced situations, which have been captured and learned –in such a way– that they can be reused in the solving of future situations. The reasoning process in the Case-Based Reasoning and Learning agent is performed by a general cycle described by the following steps:

• *Retrieving* the most similar case(s) (previous working situations) by means of some heuristic functions or distances, possibly domain dependent.
• *Adapting* or reusing the information and knowledge in that case to solve the new problem (the current working situation of the plant).
• *Evaluation* of the proposed solution. Usually, it is performed by simulation, by questioning to human oracle or by future checking of effectiveness.
• *Learning* the parts of this experience likely to be useful for future problem solving. The agent can learn both from successful solutions and from failure ones. This retaining is made by updating the Case library accordingly.

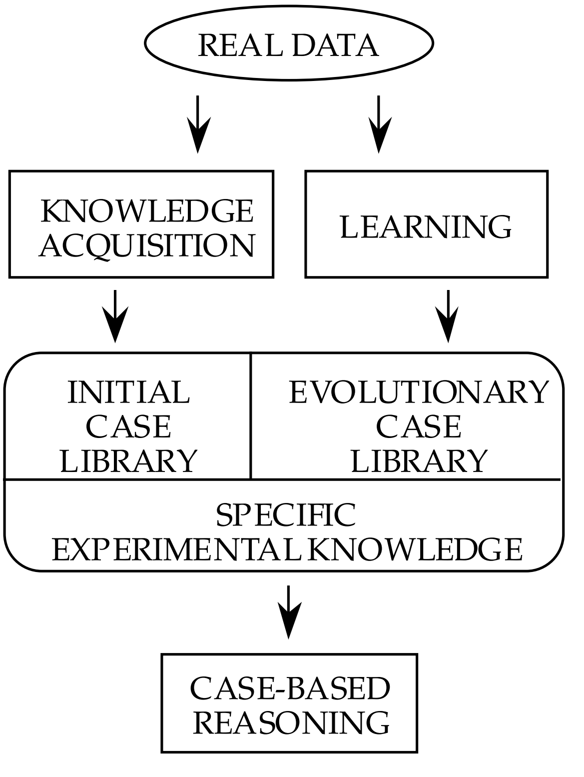

Figure 2. Specific experimental knowledge

The initial Case library is fitted with some situations obtained by Linneo$^+$ classification, from a real data stream of 521 data (days) corresponding to the period 1990-1991. Each data is described by means of the daily mean of 39 variables. That study [Sànchez *et al.*, 1994a] provided a classification of the real *specific working situations* of the concrete plant. It is interesting to notice that with these data Linneo$^+$ discovered that it exists four subtypes of *normal situations* (usually, not considered by the experts) and revealed that about 20% of the variables provided by the plant's operators were not relevant for the characterization of *situations*. The outcomming situations were:

Toxic substances loading   Normal (4)

Primary-treatment problems   Solid's shock
Plant problems   Storm
Secondary-treatment problems

So, the Case library contains a set of experienced situations of the plant (specific situations). It evolves from initial contents and captures the experimental knowledge of the concrete plant under control, learning either from its successfully solved situations (plans) or from its failed ones (objective knowledge). All this process is shown in figure 2.

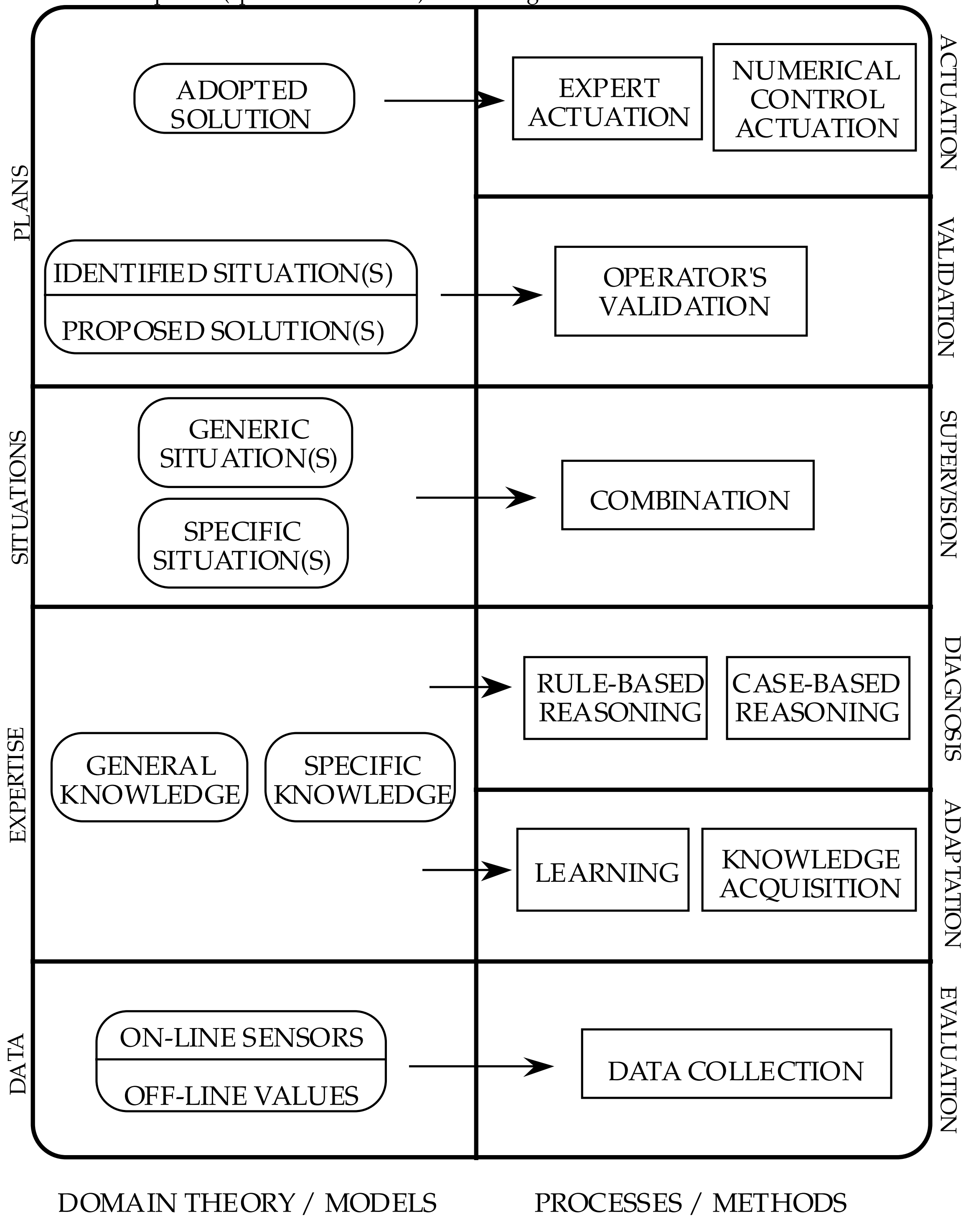

Figure 3. Integrated supervisory multi-level architecture

4 The integrated and distributed supervisory multi-level architecture

The proposed architecture (called DAI-DEPUR [Sànchez *et al.*, 1994b]) is a distributed and integrated supervisory multi-level system (see Figure 3). It is a *distributed* architecture so that is formed of several interacting subsystems (agents) that can be executed in parallel. For instance, the supervisory agent, the case-based reasoning agent, primary settler-KBS agent, biological reactor-KBS agent, etc. Distribution criteria are based on spatial and semantic distance [Bond and Gasser, 1988]. The main reason to choose a Supervisory Distributed AI System is because for WWTP there is a set of fixed *abnormal* situations as *Storm*, *Bulking*, *Toxic load*, etc., that may be treated with a predetermined plan or strategy, in a more efficiently way than other types of DAI architectures as Blackboard Systems or Contract Nets (see [Sànchez *et al.*, 1994b] for a more detailed description).

DAI-DEPUR is an *integrated* architecture because of joining in a single system several cognitive tasks as learning, reasoning, knowledge acquisition, problem solving, etc. Moreover, focusing on the *expertise level* [Steels, 1990] –that is the aim of this paper–, there is the integration of the two paradigms of knowledge modelling.

Also, the architecture is *multi-level*, and provides independence to all the levels. Taking into account the domain theory (models), it can be structured as a four-level architecture: 1) data level, 2) expertise level, 3) situations level and 4) plans level.

Data level
It is formed of on-line data gathered from sensors and off-line values provided by the operator as laboratory analysis, subjective information, etc.

Expertise level
Modelled by the two paradigms or approaches explained in this paper: general expert knowledge and specific experimental knowledge.

Situations level
The *global operating situation* of a plant is obtained by combination of its several subsystems' *local situations*.

Plans level
At this level, the identified *whole situation*, some previous solved similar situations as well as predefined (canned) plans are taken into account to propose a first solution, that has to be validated against the operator, who can modify the proposed plan. Then, an arranged plan can be executed to cope with the actual operating situation of the plant. *Plans* are a sequence of actions to be taken in order to restore the plant performance.

By the other hand, considering the processes acting over the models (methods) the architecture can be decomposed in a six-level processes or phases: 1) evaluation level, 2) adaptation level, 3) diagnosis level, 4) supervision level, 5) validation level and 6) actuation level. The system activates a new supervisory cycle at fixed intervals of time.

Evaluation process
For this purpose it is necessary to know some values for certain variables of the process. All this data can be extracted from the evolutionary Data Base, fitted either with the on-line sensors values coming from the data collecting systems or with some other features provided by the operator (like a laboratory analysis, qualitative observation, etc.).

Adaptation process
This is a process that is sometimes performed either by dynamic learning from past proposed solutions and its efficiency –that can update the Case library– or by acquiring some new knowledge from (new) experts or (new) sources through classification techniques.

Diagnosis process
In a new cycle the Supervisory agent activates the Knowledge-based agents to diagnose the state of the different subsystems of the plant by means of rule-based reasoning. At the same time in the diagnosis phase the Case-Based Reasoning and Learning agent (CBRL) is activated to retrieve similar cases recorded in the Case library. This means concurrent execution of all agents involved. Next, is updated the most similar one in order to adapt it to actual situation of the plant. For this task, needs to access to Data Base. The results are communicated to the Supervisory agent.

Supervision process
The Supervisory agent combines all information coming from the several KBS agents (*general knowledge*) and from the CBRL agent (*specific knowledge*) to infer the current global situation of the plant and the suggested actions to be taken. It sends this information to the operator through the User Interface Module.

Validation process
The system can be inquired by the operator in some ways as asking for explanations, retrieving certain values, etc. The Supervisory agent waits for the operator's validation of actions to be taken in order to update the current working state of the plant.

Actuation process
The Supervisory agent recognizes *situations* and uses the right strategy or plan, in order to keep the process controlled or if *normal situation* has been detected, then, the automatic numerical control is maintained or activated. If there are on-

line actuators, the plant can be automatically updated through the Actuator system. If not, manual operation is required.

5 Preliminary Results and Evaluation

The knowledge-based paradigm (expert knowledge) is already built-up. Knowledge-based agents are implemented using the G2 shell (in prototype status). They are incrementally validated with some real data stream taken from the plant, against the experts' opinion and some simulation studies, yielding good results [Serra *et al.*, 1993]. Also, Linneo^{+} –the knowledge acquisition module– is already implemented.

The case-based reasoning approach (experimental knowledge) is under development. The Case library is being implemented as a prioritized discrimination network, where the priority of node-attributes is obtained from experts' opinion and from an inductive learning method (such as ID3, etc.). Cases retrieving and adding tasks are designed and, now, we are working on adaptation and evaluation steps. This agent is implemented in Lisp, such other agents (supervisory, etc.).

Data collection module is also developed. It gathers the values from status of turbines, pumps, automatic grids, etc., from the control panel of the plant under supervision, with 6 PLC's (SISTEL 8512). These 96 digital signals are transmitted through RS-422 to the monitoring computer (PC) for evaluation process. It also receives 9 analogical signals (inflow, wasting flow, recirculation flow, biogas produced, DO-line-1, pH, temperature-digester-1 and temperature-digester-2) converted by an AD/DA card. It will be connected to the main computer (SUN Sparc station) where are running all the other agents and processes.

After the validation on the simulation of the plant, the prototype system is currently being validated –in an incremental way– in a real plant under contract between our universities and "Junta de Sanejament de la Generalitat de Catalunya" as the responsible organism of wastewater treatment plants management in Catalonia. Until now, the results obtained with the system are promising.

6 Summary

From what has been previously described, one can state that the combination of both paradigms at the *expertise level*, let the system to model the subjective knowledge (supplied by the experts) as well as the objective knowledge (supplied by the real operation of the concrete plant under control)[1]. This integration presents some advantages that are the addition of the own ones from knowledge-based reasoning, from case-based reasoning, from dynamic learning and from semi-automatic knowledge acquisition, that are the methods acting at the expertise level:

• It makes possible to *reason in a poor understood and ill-structured domain*, where other kinds of reasoning like model-based reasoning or algorithmic reasoning could not be possible.
• Let the system to learn from previously solved problems and to adapt the available experimental knowledge over the domain (*dynamic learning environment*).
• Overcoming the *brittleness* of KBS' in coping with unforeseen situations (not previously considered by the general expert knowledge), trying to solve them by means of the most closely situation in the Case library.
• Capturing the knowledge provided by the experts (*knowledge acquisition*) which is very important –although subjective– to get a central corpus of knowledge about the domain.
• Dealing either with prototypical situations (*general knowledge*) or with idiosyncratic or exceptional ones (*specific knowledge*).
• Due to the dynamic learning environment, the system is able to adapt itself to different wastewater treatment plants, *making the system to be exportable* to any plant with some minor changes. It is only needed to fill the Case library with an initial set of specific cases (operating situations of the concrete WWTP), which can be obtained semi-automatically from real operational data.

All these facts make it more powerful than other single technologies applied to wastewater treatment plants as knowledge-based approaches [Lapointe *et al.*, 1989; Maeda, 1989], statistical process control techniques [Novotny *et al.*, 1990], fuzzy controller methods [Czoagala and Rawlik, 1989; Alex *et al.*, 1994], etc., as well as to other complex ill-structured domains. With this approach, the plant can be controlled in *normal* situations (mathematical control), in *abnormal usual* situations (expert control) and in *abnormal unusual* situations (experimental control).

Acknowledgements

[1]So that it is commonly known –and has been showed out in the performed study [Sànchez *et al.*, 1994a]– that not all the considered situations by the experts occur in the practice, and vice versa, certain situations not taken into account by the experts can occur in the WWTP.

Currently, it is been partially supported by the "Junta de Sanejament de la Generalitat de Catalunya" and the Spanish CICyT project ROB94-679 and EEC project VIM ERBCHRXCT 930401. Also the authors wish to acknowledge the cooperation of Ricard Tomàs, manager of the Manresa's wastewater treatment plant.

References

[Alex *et al.*, 1994] J. Alex, U. Jumar and R. Tschepetzki. A Fuzzy Controller for Activated Sludge Wastewater Plants. In *Procc. of the 2nd IFAC/IFIP/IMACS Int. Symp. on Artificial Intelligence in Real Time Control (AIRTC'94)*, pages 75-80, València, October, 1994.

[Bond and Gasser, 1988] Alan H. Bond and Les Gasser (editors). *Readings in Distributed Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[Béjar, 1995] Javier Béjar. Knowledge acquisition in ill-structured domains. Ph. D. Thesis. Dept. de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya, 1995. (In Spanish).

[Czoagala and Rawlik, 1989] E. Czoagala and T. Rawlik. Modelling of a Fuzzy Controller with application to the Control of Biological Processes. *Fuzzy Sets and Systems,* 31:13-22, 1989.

[Kolodner, 1993] Janet Kolodner. *Case-Based Reasoning*. Morgan Kaufmann, 1993.

[Lapointe *et al.*, 1989] J. Lapointe, B. Marcos, M. Veillette, G. Laflamme and M. Dumontier. Bioexpert – an Expert System for Wastewater Treatment Process Diagnosis. *Computers & Chemical Engineering*, 13(6):619-630, 1989.

[Maeda, 1989] Kazuo Maeda. A Knowledge-based system for the wastewater treatment plant. *Future Generation Computer Systems*, 5:29-32, North Holland, 1989.

[Novotny *et al.*, 1990] V. Novotny, H. Jones, X. Feng and A. G. Capodaglio. Time Series Analysis Models of Activated Sludge Plants. *Water Science & Technology*, 23(4-6):1107-1116, 1990.

[Patry and Chapman, 1989] Gilles G. Patry and David Chapman (editors). *Dynamic Modelling and Expert Systems in Wastewater Engineering*. Chelsea, MI. Lewis Publishers, 1989.

[Plaza *et al.*, 1993] Enric Plaza, Agnar Aamodt, Ashwin Ram, Walter Van de Velde and Maarten Van Someren. Integrated Learning Architectures. In *Procc. of the European Conference on Machine Learning (ECML-93)*, LNAI-667, pages 429-441, Springer-Verlag, 1993.

[Riesbeck and Schank, 1989] Christopher K. Riesbeck and Roger C. Schank. *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates Publishers, 1989.

[Robusté, 1990] Jordi Robusté. Modelling and identification of the activated sludge process. Ph. D. Thesis. Dept. de Química. Universitat Autònoma de Barcelona, 1990. (In Catalan).

[Sànchez *et al.*, 1994a] Miquel Sànchez, Ulises Cortés Joan de Gràcia, Javier Lafuente and Manel Poch. Concept Formation in WWTP by means of Classification Techniques: a Compared Study. Submitted for publication. August, 1994.

[Sànchez *et al.*, 1994b] Miquel Sànchez, Ignasi R.-Roda, Javier Lafuente, Ulises Cortés and Manel Poch. DAI-DEPUR Architecture: Distributed Agents for Real-Time WWTP Supervision and Control. In *Procc. of the 2nd IFAC/IFIP/IMACS Int. Symp. on Artificial Intelligence in Real Time Control (AIRTC'94)*, pages 179-184, València, October, 1994.

[Schank and Slade, 1991] Roger C. Schank and Stephen B. Slade. The future of Artificial Intelligence: learning from experience. *Applied Artificial Intelligence*, 5:97-107. Hemisphere Publishing Corporation, 1991.

[Serra *et al.*, 1993] Pau Serra, Miquel Sànchez, Javier Lafuente, Ulises Cortés and Manel Poch. ISCWAP: a knowledge-based system for supervising activated sludge processes. Submitted for publication, 1993.

[Serra *et al.*, 1994] Pau Serra, Miquel Sànchez, Javier Lafuente, Ulises Cortés and Manel Poch. DEPUR: a knowledge based tool for wastewater treatment plants. *Engineering Applications of Artificial Intelligence*, 7(1):23-30, 1994.

[Steels, 1990] Luc Steels. Components of expertise. *AI Magazine*, 11 (2):28-49, 1990.

[Stephanopoulos and Stephanopoulos, 1986] George Stephanopoulos and Gregory Stephanopoulos. Artificial Intelligence in the

Development and Design of Biochemical Processes. *Trends in Biotechnology*, pages 241-249, September, 1986.

[Stephanopoulos, 1990] G. Stephanopoulos. Artificial Intelligence in Process Engineering: current State and future Trends. *Computers & Chemical Engineering*, 14:1259-1270, 1990.

[Venkatasubramanian, 1994] Venkat Venkatasubramanian. Towards Integrated Process Supervision: Current Status and Future Directions. In *Procc. of 2nd IFAC Workshop on Computer Software Structures Integrating AI/KBS Systems in Process Control*, pages 9-21, Lund, August, 1994.