Integration of a rule-based expert system, a case-based reasoner and an ontological knowledge-base in the wastewater domain

Luigi Ceccaroni[1]

Abstract. We present an environmental decision-support system integrating a rule-based expert system, a case-based reasoner and an ontological knowledge-base. This system is able to model the information about a wastewater treatment process through the definition of the basic terms and relations comprising the vocabulary of the wastewater treatment area. Furthermore, this management system optimizes the operation of wastewater treatment by a *more reliable management* and making easier its self-*portability*.

1 INTRODUCTION

The general issues we would like to address in the paper are:

* optimizing wastewater treatment operation by a *more reliable management* and
* making easier the *portability* of the management system.

1.1 Physical environment

Wastewater purification Contamination levels of waters constantly increase due basically to industrial development and to the increase of population density in certain zones. Wastewaters, either industrial or urban, have to be decontaminated until an adequate level, so that they could be poured to the surrounding hydric medium without causing problems of environmental deterioration. For that, there are *wastewater treatment plants* (WWTPs) with techniques of physical-chemical and biological treatment.

Every treatment process carries, in greater or minor measure, economic and environmental costs as for that it generates another kind of waste that needs, in turn, of other elimination techniques. In this sense, for the case of wastewaters, the use of biological processes of purification is favored over the physical-chemical ones. Biological processes, in general, almost do not consume reagents, they are more efficient, they do not generate either gases or noxious sludges, and they are responsible of a higher production of sludges that can be used (not always) in other productive processes (e.g. as fuel, fertilizer and filling material).

Wastewater treatment plants with biological processes are the physical environment modeled by our system. The general operation of a WWTP always includes various internal pre-designed standard units, whose sub-operation is optimized to accomplish a single task[2] [13]. Each sub-operation usually has effects on other downstream treatment processes, and tradeoffs between increasing the efficiency of one process or another are necessary, taking into account major constraints such as water characteristics, effluent quality and costs of each operation.

1.2 Software environment

The system we propose (named DAI-DEPUR+) receives on-line inputs[3] from sensors all over the WWTP as well as off-line inputs from the WWTP laboratories and human operators. The system uses its internal knowledge-bases[4] and inference[5] mechanisms to process and understand this information, to diagnose the ongoing WWTP-state, and to predict the evolution of that WWTP state. Eventually, the output of the system is represented by statements about actions to be taken, or statements to support human decisions in future actuations, or direct control signals to WWTP devices in order to maintain the plant working correctly.

In case of diagnosis impasse, DAI-DEPUR+, before turning to the plant manager, will try to solve the problem exploiting the connection, in the ontology, between data and states of the WWTP.

1.3 Motivations

The process of wastewater treatment is so *complex* that it is difficult to develop a reliable supervisory technology based only on a chemical-engineering classic-control approach. Nowadays the use of Artificial-Intelligence (AI) systems seems to be necessary in order to obtain better results in wastewater management.

Rule-based expert systems (one of the broadly applied paradigms of AI) proved able to cope with some known difficulties and to face several WWTP-domain problems, even if they are not the definitive solution to the treatment problem as a whole. On the other hand, large, multifunctional, available ontologies would significantly improve current expert systems and tutoring systems because they contain the broad knowledge of a domain required to perform multiple

[1] Universitat Politecnica de Catalunya, Departament de Llenguatges i Sistemes Informatics, Campus Nord, Modul C6, Jordi Girona 3, 08034 Barcelona, Spain, email: luigic@lsi.upc.es

[2] A task in this context is generally the removal / remediation of contaminant substances or pathogenic microorganisms.

[3] Input/output devices are any of various devices used to enter information and instructions into the DAI-DEPUR+ system for storage or processing and to deliver the processed data to a human operator or, in some cases, a machine controlled by the system. Such devices comprise sensors and effectors. Apparatus of this kind with direct connection to DAI-DEPUR+'s central processing unit is said to be on-line; peripheral equipment working independently of it is termed off-line.

[4] Both static (rule-based), dynamic (case-based) and ontological knowledge bases.

[5] Inference is the process of drawing conclusions about a particular parameter of the domain.

tasks and to explain domain knowledge from multiple viewpoints. A lot of ontologies are nowadays being built in many research centers around the world, but a few of them are specialized in biology or ecology, let alone wastewater management and WWTP microbiology.

A great improvement in addressing this kind of problems can come from the *integration* of different modeling and reasoning systems, such as specific *ontologies*, *rule-based reasoning*, *case-based reasoning* and *reactive planning*. The architecture we adopt (DAI-DEPUR+) integrates two knowledge-based systems and an ontology, and is flexible enough to deal with the complexity of the wastewater treatment process, given an adequate amount and kind of data. In the DAI-DEPUR+ architecture, with the embedded ontology for the wastewater treatment, the representation of a deeper knowledge of the domain is permitted and the evolution of WWTP-microorganism communities can be taken into account. In this way, the management of biological problems arising in a treatment plant is more effective.

For the first time, by the integration of an ontology with two knowledge-based systems, it will be possible to capture, understand and describe the knowledge about the whole physical, chemical and microbiological environment of a wastewater treatment plant. The basic terms and axioms of the ontology will entail a model of wastewater domain and a classification of the microorganisms according to known biological taxonomy, and will include a complete description of the microorganisms themselves (physical aspect, abundance and behavior in the treatment plant). The relations of the ontology will include, besides the classic hierarchical affiliations, all the interesting bindings among microorganisms and between them and the state of the plant (diagnosis potential).

1.4 General overview

In this paper we present a decision support system for the supervision of wastewater treatment plants, which is part of the knowledge and technology needed for the rational management of water resources.

We start by describing (in section 2) related work on the environmental-domain study and the AI techniques (including ontologies) related to the creation of environmental decision-support systems. In section 3 we explain how the decision support system (DAI-DEPUR+) has been designed and we include a description of its layered architecture. Eventually, in section 4 the contributions of this work are summarized and discussed.

2 WASTEWATER DOMAIN AND AI TECHNIQUES

2.1 Wastewater treatment process

In this section we describe the general treatment process, its possible variations, and a wastewater description from a physical, chemical and biological point of view.

Wastewater treatment process The wastewater treatment process is part of the water cycle and, as such, it has a direct relation with other water systems or reservoirs. Wastewater treatment plants (WWTPs) receive water from the anthropic system of sewers, they somehow process it, and finally they deliver this water to a natural reservoir. The wastewater processing is what we care about, but we cannot forget the two other closest components of the global water cycle just mentioned (sewers, and river or sea).

It is on the basis of the quantity and quality of water to be treated that WWTPs are built, taking into account the possible fluctuations in the inflow. These fluctuations can be very important where the sewerage system is not very developed and therefore it is not able to damp down inflow peaks towards the plant.

The main objectives in wastewater-treatment research are:

- knowing better the relevant characteristics of the wastewater,
- refraining the contaminated water from reaching the natural environment.

The fact is that continuously increasing economic and cultural pressures on freshwater resources, including pollution and excessive use, are causing threats which are augmenting costs and multiplying conflicts among different users of this strategic resource. These pressures can also impair the natural regenerative functions of the ecosystems in the water cycle. Two of the main challenges in the area of general water-management are to protect the water bodies and to provide high quality water in sufficient quantity at affordable costs. In order to achieve these goals, multidisciplinary research-efforts and actions are necessary. The very existence of WWTPs and the research for improving them goes in this direction and constitutes an essential element for an integrated sustainable management of water resources. The objectives of such sustainable management are to develop technologies to prevent and treat pollution of water, to purify water, to use and re-use it rationally, to enhance efficient treatment of wastewater and to minimize environmental impacts from wastewater treatment (including the prevention of potential health hazards).

2.1.1 General wastewater characterization

Urban wastewater can be characterized in accordance with the presence of different kinds of dumping, such as domestic, commercial or industrial ones. Another important feature is the presence of pathogenic organisms, which can prejudice a possible alternative reuse of treated water, such as irrigation.

There are substantially two components in wastewater: human metabolic waste and discarded material. While the first component is almost changeless in nature (as it is dependent on human metabolism), the second one depends on many parameters, such as standard of living, local habits and country.

2.1.2 Physical components

Total solids can be distinguished in suspended (sedimentable or not), colloidal and dissolved, and contain organic and inorganic portions. The size of the solids that are present in wastewater influences the sedimentation, adsorption, diffusion, mass transfer and biochemical reactions. The *temperature* of wastewater depends on the typology of dumping and on the permanence time in the sewers. Except for summer months, it is higher than environment temperature, due to the presence of warm water dumping from kitchens and bathrooms. The importance of wastewater temperature is bound to the biological activity of purification in treatment plants. At more than 40°C nitrification halts and temperatures higher than 50°C block aerobic digestion. Temperatures lower than 15°C inhibit the anaerobic methanogenic process, while at 5°C the nitrificant autotrophic flora stops its activity and at 2°C also the heterotrophic flora become ineffective. Wastewater *color* is strictly correlated to its age, its septic conditions and to the presence of industrial dumping. The *odor* is associated to putrescence and decomposition degree of organic matter, and to the presence of particular industrial wastewater.

2.1.3 Chemical characteristics

Here, a brief description of organic and inorganic chemical descriptors of wastewater is given.

In general, *organic matter* is rapidly biodegraded, but part of it is not and moreover is toxic for many WWTP microorganisms. To evaluate the content of organic matter, the *biochemical oxygen demand*[6] and the *chemical oxygen demand* (COD) are determined.

The majority of toxic effects on WWTP-microorganisms' growth are attributable to *inorganic matter*, such as *heavy metals*, and to its interaction with other wastewater materials.

The *nitrogen* found in wastewater is of five prevalent kinds: organic nitrogen (in vegetal and animal proteins), ammoniacal nitrogen, nitrites, nitrates and elemental gaseous nitrogen. Ammoniacal nitrogen is produced during the decomposition / hydrolysis of organic nitrogen and can come from the bacterial reduction of nitrites or directly from industrial dumping.

The main kinds of *phosphorus* existing in wastewater are: salts of orthophosphoric acid, polyphosphates and organic phosphorus. In urban wastewater, in general, all kinds of phosphorus are present, while, after a biological treatment, there are generally only orthophosphates.

Sulfur is present in the form of sulfates or sulfides. Sulfates can be reduced to sulfides by sulfate-reducer bacteria in anaerobic conditions. Sulfites constitute a culture medium for several species of aerobic bacteria able to create sulfuric acid, which can cause corrosion problems.

Chlorides have metabolic human origin (as they are contained in urine in an amount equal to 1%) or are due to industrial-water contribution.

Some *heavy metals* in wastewater are necessary in minimum amounts as microelements for WWTP microorganisms and for aquatic life, but they are poisonous in high concentrations.

2.1.4 Biological components

A basic knowledge about the most common natural organisms that can be found in wastewater is also necessary to control the treatment process. Some of these organisms are essential for certain pollution-removal treatments, such as activated sludge. The majority of pathogenic organisms are part of human intestinal bacterial flora and they cannot survive for a long time in wastewater. In general, most of the organisms of human origin are banal saprophytic bacteria, that is organic-matter demolishers; they are not pathogenic and can enter biological processes without any problem [12].

2.1.5 Wastewater treatment plants

In a wastewater treatment plant (WWTP), the main goal is to reduce the level of pollution of the inflow water, that is to remove, within certain limits (depending on local legislation), abnormal amounts of pollutants in the water prior to its discharge to the natural environment. This can be done in a number of different ways, corresponding to different kinds of WWTP. The most widespread classes of WWTP are:

- plants with only physical-chemical treatment;
- plants with additional biological reactor (for better organic matter removal), which can be of two main sub-type, depending on the sort of growth of microorganisms [5]:
  - suspended growth: with the microorganisms mixed with the wastewater and dispersed in the form of free cells or of bioflocks (activated sludge reactors).
  - attached growth: with the microorganisms anchored, in the form of biofilm, to inert surfaces (biological-film reactors).

The work of the paper focuses on WWTPs with activated sludge, which is now the most common case in the European Union.

2.2 Rule-based expert systems

Rule-Based Expert Systems (RBESs) are advanced computer programs which emulate, or try to, the human reasoning and problem-solving capabilities, using the same knowledge sources, within a particular discipline [23] [26] [9] . RBESs always possess certain heuristics that form the static knowledge-base, and some inference and search processes. The problems addressed with RBESs are very complex and related to specific domains, and they would usually need a very expert human (i.e., a great amount of knowledge) to be solved[7]. A few examples of real-world general applications of RBESs are the following ones:

- decision support for natural resources management [18],
- data management in forestry [31].
- petrochemical-plant control [1],
- dynamic-process monitoring and diagnosis [20],
- WWTP time-series analysis [33],
- control of sun-powered systems [38].

The main components of RBESs are: static knowledge-base (or long-time memory), data base (or working memory or short-time memory), inference engine, user interface, auto-explanation module, strategy module, knowledge-engineer interface and on-line sensor/effectors interface.

Typically, the knowledge contained in the historical data is encoded in the static knowledge-base in the form of rules or axioms, via a knowledge-acquisition process. The rules allow the system to deduce new results from an initial set of data (premises). A rule is basically represented by the following code:

IF *conditions* THEN *actions*

The reasoning method (inference engine) may use forward chaining, backward chaining or a combination of both of them. Forward-chaining reasoning (deduction) starts from the input data towards the final conclusions, deducing new facts from previous ones. Backward-chaining reasoning (induction) is guided by the conclusions towards the input data (commonly provided by the user).

Thanks to their characteristics, RBESs have been widely and successfully applied to environment management, supervision and control [11] [36] [30] [14] [41].

2.3 Experiential knowledge and *case-based reasoning* (CBR)

CBR is both a paradigm for computer-based problem solvers and a model of human cognition. The central idea is that the problem solver reuses the solution from some past case to solve a current problem.

[6] The BOD represents the amount of oxygen needed by bacteria to degrade the organic matter and it is function of the organic matter concentration and of the degradation rate.

[7] It may even happen that the RBES algorithmic-power could do some special tasks that the human one (the mind) cannot do in the great majority of the cases.

2.3.1 CBR as a computer program paradigm

As a paradigm for computer-based problem solvers, one of the advantages of CBR systems is that they improve their performance, becoming more efficient, by recalling old solutions given to similar problems and adapting them to fit the new problems. In this way they do not have to solve new problems from scratch[8]. The memorization of past problems / episodes is integrated with the problem-solving process, which thus requires the access to past experience to improve the system's performance. Additionally, case-based reasoners become more competent during their functioning over time, so that they can derive better solutions when faced with equally or less familiar situations because they do not repeat the same mistakes (learning process). The basic steps in CBR are:

1. Introducing a new problem (or situation) into the system.
2. Retrieving a past case (a problem and solution) that resembles the current problem. Past cases reside in case memory. Case memory is a database that contains rich descriptions of prior cases stored as units. Retrieving a past case involves determining what features of a problem should be considered when looking for similar cases and how to measure degrees of similarity. These are referred to as the Indexing Problem and the Similarity Assessment Problem.
3. Adapting the past solution to the current situation. Although the past case is similar to the current one, it may not be identical. If not, the past solution may have to be adjusted slightly to account for differences between the two problems. This step is called Case Adaptation.
4. Applying the adapted solution and evaluating the results.
5. Updating case memory. If the adapted solution works, a new case (composed of the problem just solved and the solution used) can be formed (direct learning). If the solution at first fails, but can be repaired so the failure is avoided, the new case is composed of the problem just solved and the repaired solution. This new case is stored in case memory so that the new solution will be available for retrieval during future problem solving. In this way, the system becomes more competent as it gains experience. Updating case memory includes deleting cases (forgetting), too. This step is also part of the Indexing Problem.

Not all case-based problem solvers use all of the steps. In some, there is no adaptation step; the retrieved solution is already known to be good enough without adaptation. In others, there is no memory update step; the case memory is mature and provides adequate coverage for problems in the domain.

2.3.2 CBR and wastewater environment

In the WWTP domain, CBR has been used for designing more suitable operations to treat a set of input contaminants [29] and for supervision [36] [37]. In this context, the cases stored in the case library are real WWTP operating states, which are learned in such a way that it is possible to reemploy them to solve future tasks. A case incorporates the following set of features: an identifier, the situation description, the situation diagnosis, the action plan, the derivation (from where the case has been taken / adapted), the solution result (success / failure), a utility measure, a distance / similarity value.

2.3.3 CBR's problems

In general, case-based reasoning proved to be a good choice for experiential-knowledge (specific-knowledge) management. But CBR has the basic problem that it cannot work alone if there is no available experience, such as in the case of the initial running period of a treatment plant. It has to be combined, for instance, with a rule-based or an ontology-based system (general-knowledge managers) so that it can work as a reasoning component in the overall control and supervision of WWTPs. An integration of different AI methods is needed, that includes the management of qualitative information (e.g. microbiological descriptors, in the case of wastewater treatment), experts' intelligence and experiential knowledge.

[8] Non-blind problem-solving strategy.

2.4 Ontologies

2.4.1 AI definitions

AI literature is full of different definitions of the term *ontology*. Each community seems to adopt its own interpretation according to the use and purposes that the ontologies are intended to serve within that community.

- One of the early definitions: 'An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary.' [32]
- A widely used definition (Gruber): 'An ontology is an explicit specification of a conceptualization.' [24]
- An elaboration of Gruber's definition: 'Ontologies are defined as a formal specification of a shared conceptualization.' [6]

2.4.2 Ontological knowledge-bases

Knowledge bases (KBs) run through a spectrum from simple collections of frequently-asked questions (FAQs) to complex systems powered by AI engines. Historically, the term *knowledge base* refers to a base of expert information and answers to common questions. By processing its knowledge base using rules called heuristics, an expert system can respond to a series of questions and choices, and solve a problem as though the user were dealing with a human expert in a particular field. Today, the term *knowledge base* has developed at least two second meanings:

- One in the context of the world wide web. In this domain, a knowledge base is simply a base of technical information or answers to common problems, often related to a particular system or product. These web knowledge-bases (WKBs) may be provided as a customer service on a corporate web site, or they may be developed by knowledge engineers for knowledge workers within an institution or company. While some of these knowledge bases are used by expert systems or other AI systems to solve problems, most are just part of simpler search engines, like those generally used to search the Web.
- One in the context of AI. In this paper we refer to KBs only as:
  - *ontological knowledge-bases* (OKBs): in the domain of AI-ontologies, a KB is the computer-readable translation of an ontology; these OKBs are sometimes part of a more general expert system;
  - knowledge bases of rule-based expert systems or *static knowledge-bases* (SKBs);
  - knowledge bases of case-based reasoners or *dynamic knowledge-bases* (DKBs).

  In AI, KBs were born to help in knowledge reuse and sharing: 'reuse' means building new applications assembling components

already built, while 'sharing' occurs when different applications use the same resources. Reuse and sharing present the following advantage: need for less money, less time and less resources.

2.4.3 Knowledge sharing and reuse

When sharing knowledge, it is possible to come across problems relative to:

- the conceptualization method [22],
- the shared vocabulary (e.g., libraries of ontologies),
- the format to exchange knowledge (e.g., KIF (Knowledge Interchange Format)), and
- the specific communication protocol (e.g., KQML (Knowledge Query Manipulation Language) external interface).

When reusing knowledge, the most common problems concern:

- the heterogeneity of knowledge-representation formalisms and of the implementation languages (worked out by translators),
- the lexicon,
- the semantics,
- synonyms and hidden assumptions (worked out by the very ontologies), and
- the loss of common-sense knowledge (addressed by an integration of various AI paradigms, such as ontologies, natural language processing and machine learning, with cognitive science) [21].

2.4.4 Ontology development: from art to understood engineering process

Even if it is now widely recognized that constructing ontologies, or domain models, is an important step in the development of KBSs, what is lacking is a clear understanding of *how* to build ontologies. However, there exists a small but growing number of methodologies that specifically address the issue of the development and maintenance of ontologies. In this section we present, among the projects which go in the direction of providing these methodologies, the one which is most related to the ontology of the DAI-DEPUR+ system: the Ontolingua Project. For a comprehensive survey of the work which has been done so far, see [27].

Ontolingua The guides for the use of the Ontolingua Ontology Server [15] [16] [19] contain advice on developing, browsing, maintaining and sharing ontologies through the Server. The Ontolingua language is based on the syntax and semantics of KIF. One of the main benefits in using the Ontolingua server is the access it provides to a library of previously defined ontologies. This library extends as developers add new ontologies to the repository.

Ontology construction in Ontolingua is based on the principle of modular development. Ontologies from the library can be re-used in four different ways:

1. inclusion: *ontology A* is explicitly included in *ontology B*. The vocabulary of *ontology A* is translated into the vocabulary of *ontology B*. This translation is applied to the axioms of *ontology A*, too, and the translated axioms are added to *ontology B* [16]. Multiple inclusion is supported.
2. polymorphic refinement: a definition from an ontology is included in another ontology and refined. For example, the Biological-Living-Object class, defined in UpperCyc ontology, can be included in WaWO ontology and extended to admit Bacteria, and included in *ontology B* and extended to admit aliens.
3. restriction: a restricted (by axioms) version of one ontology is included in another.
4. cyclic inclusion: as ontology inclusion is transitive, situations such as the following are allowed, even if not recommended: *ontology A* is included in *ontology B*, *ontology B* is included in *ontology C* and *ontology C* is included in *ontology A*.

These distinctions are very useful in the re-use of ontologies, but the specification of the relationships among ontologies is probably not complete [27]. Ontolingua is the *de facto* standard means of implementing ontologies although a more comprehensive methodology needs to be used in conjunction with the Server.

One of the main efforts of the Ontolingua project concerns the representation of uncertain knowledge within an ontology. The *Ontolingua representation language* resulting from this work enables ontologies to contain richly textured descriptions that include uncertainty, are structured into multiple views and abstractions, and are expressed in a generic representation formalism optimized for reuse. The Ontolingua language uses the Knowledge Interchange Format (KIF) as a core. It is a computer-interpretable description language and enables easy on-line collaborative construction of ontologies [24]. (http://ontolingua.stanford.edu/)

With respect to ontology editors, there are a number of more or less generic editors to create and manage ontologies. The Stanford Ontolingua Ontology Editor (Stanford KSL Network Services[9]) is the most standard editor to create ontologies.

2.4.5 Ontologies and the environment

Environmental ontologies are just instantiations of the general ontology concept which assist in understanding a domain related to the natural environment and in modeling the processes involved. No ontology application exists yet in the field of WWTPs and no ontology modeling the evolution of microbiological systems has been defined. We think that the representational power of ontologies can be exploited to deepen the knowledge about the microorganisms of WWTP activated-sludge and the wastewater domain in general, and can be integrated together with other reasoning methods to better the whole supervision of WWTPs.

2.5 *Environmental decision support systems* (EDSSs)

When dealing with problems which have a negative impact on the environment, there are questions that managers in the public or private domain have not the time or the inclination to consider and, furthermore, they may not have sufficient knowledge of environmental issues to carry out an assessment in anything other than an entirely 'ad hoc' manner. Thus, EDSSs are called for.

An EDSS is an integrated KBS, applied to an environmental issue, that reduces the time in which decisions are made and improves the consistency and quality of those decisions [25]. In this section, we discuss which features an EDSS should include.

An EDSS should include the following features:

- The ability to assist the user during problem formulation, that is, deciding which objectives need to be reached, and when and how the different available tools have to be applied.
- A structured framework, which draws information from the user and the environmental system about domain-characteristics and

[9] http://www-ksl-svc.stanford.edu:5915/&service=frame-editor

processes in a logical manner. This framework, besides acquiring the domain knowledge, has to be able to organize and represent it.

- Specific knowledge-bases pertinent to the type of domain being considered or to the process being carried out at the site. These knowledge bases contain data on environmental parameters and processes that are relevant to the domain (e.g. what processes are required to manufacture a particular product; what toxic materials are used in the processes; which kinds of physical, chemical and biological samples need to be collected; which is the relative importance of the features in play; which are the requirements of the local legislation).
- A general environmental knowledge which is used to deduce the relative significance of different environmental impacts given appropriate data about the specific domain and processes.
- A module to present the analysis' results in a user-friendly manner.
- The ability to assist the user during the interpretation of the results and the selection of the solution. This can be done by identifying the significant features of the analysis' results and evaluating their impact with respect to the task being performed.

3 THE DAI-DEPUR+ ENVIRONMENTAL DECISION-SUPPORT SYSTEM

In this section we describe the DAI-DEPUR+ environmental decision-support system (EDSS) for wastewater treatment plants. The DAI-DEPUR+ system, as explained in detail below, includes an *ontology* which helps to model the wastewater treatment process, paying a special attention to the management of the qualitative knowledge, that is the environmental information on microorganism presence. As well as helping to model the domain, the ontology adds new capabilities to the EDSS, such as support of causal reasoning, prediction, and semi-automatic generation of a static KB.

The DAI-DEPUR+ system has an architecture in which several artificial intelligence techniques integrate and operate in real time. Of particular interest is the integration of the ontology for the representation of the wastewater treatment process. The DAI-DEPUR+ system is built to manage specific WWTPs, but the ontological representation of the domain will make easier its *portability* towards other WWTPs and other domains. The DAI-DEPUR+ system derives from the DAI-DEPUR system [36]. It is its direct evolution and is constantly under development in relation with the research of the Knowledge Engineering and Machine Learning (KEML) group at UPC. Indeed, the DAI-DEPUR+ system aims to go a step further in completing the comprehension of WWTP-microorganisms through the use of the ontology and exploiting the data on activated sludge.

In this section we explain the architecture of DAI-DEPUR+ and its 3 layers: perception, diagnosis and decision support.

3.1 Architecture

The architecture of the system has a modular design, to improve modifiability, understandability and reliability. It basically follows a standard vertical decomposition approach[10]: a division is made into many specialized subsystems, such as perception, diagnosis, modeling, planning, execution and effector-control modules.

Fig.1 contains a block diagram of the top-level decomposition of this architecture. The system receives raw data from the sensors and the laboratory, and emits commands to the sensors and effectors. The action component takes the output of the perception component as input and it is the one which generates commands to both the sensors and effectors. This differs from many other systems, in which the control of the sensors is the responsibility of the perception component.

[10] For the definition and application of horizontal and vertical decomposition, see [8] and [28].

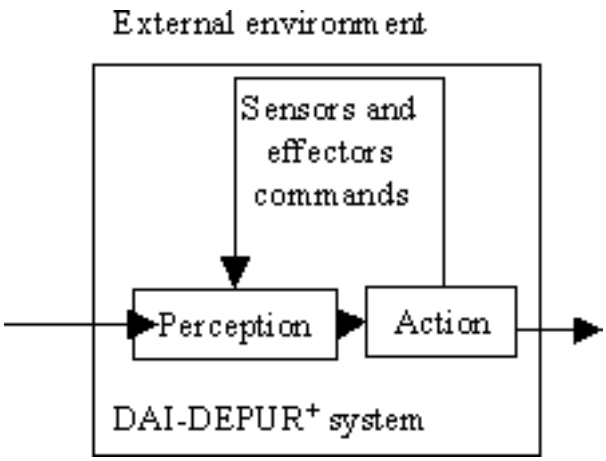

Figure 1. Top-level decomposition of DAI-DEPUR+.

Excepting cases of failure, there is a continuous sensory data stream from all sensors, which goes directly into the perception component, along with the results of laboratory analyses and the commands that were last sent to the effectors.

The detailed architecture of DAI-DEPUR+ is schematized in Fig.2 and its action model is the following one:

- perception: data gathering and knowledge acquisition,
- diagnosis: reasoning,
- decision support: prediction, evaluation of alternative scenarios, advising, actuation and supervision.

3.2 Perception layer

The DAI-DEPUR+ system operates in a domain which physically consists of a wastewater treatment plant. In particular, all the physical, chemical and biological measurements are gathered in treatment plants located in Catalunya. Some parameters are measured on-line by sensors, while other ones are measured off-line in laboratories.

3.2.1 Awareness

The time scales of the treatment processes are long, so that the perception and the supervision decisions easily fit between sampling points.

Many decision support systems simply "close their eyes" while a time-consuming subsystem, such as a planner or a reasoner, is invoked and the penalty for such unawareness is that perceptual inputs are either lost or stacked up for later processing. This is not the case in DAI-DEPUR+ because the WWTP environment is very slowly evolving compared to the speed of the reasoning of the decision support system: even if a WWTP is a truly dynamic domain, it never changes to such extent that the results of relatively long calculation would no longer be useful. If something happens that requires "immediate" action on the part of the system, DAI-DEPUR+ is always aware of it.

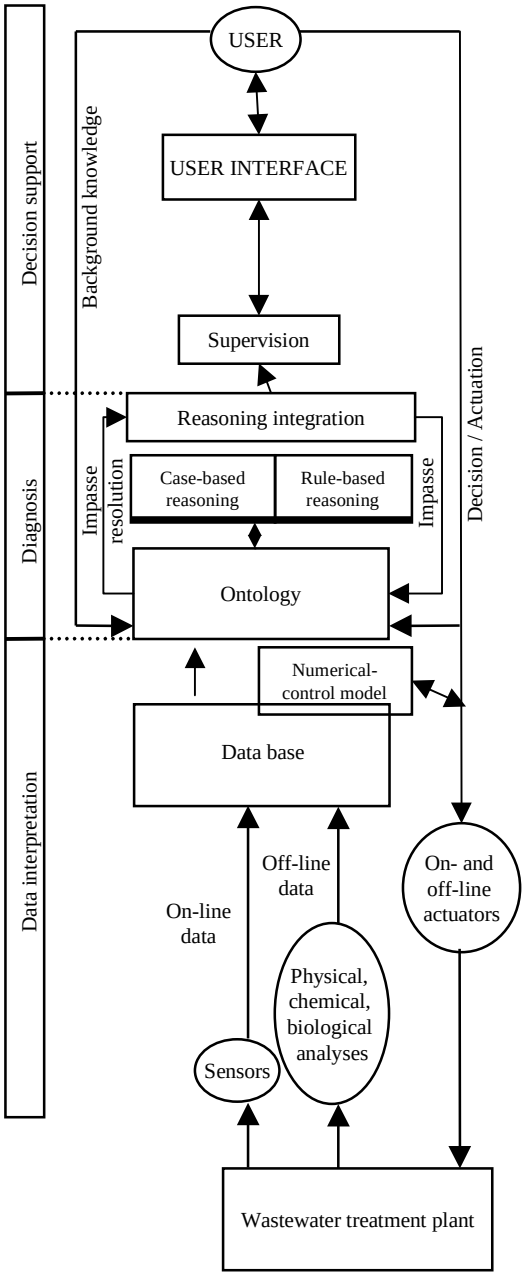

Figure 2. A view of DAI-DEPUR+ architecture.

3.2.2 Temporal integration

In a WWTP, sample intervals range from a few seconds to a few days. Our approach to the temporal integration of a number of processes that work at different rates is to define a constant minimum-cycle time for the entire system. This time is equal to 1 hour and at each tick of this time the inputs are read or calculated, some computation is done and the outputs are set (by the action component of DAI-DEPUR+). If a process, such as a laboratory analysis, cannot complete (or even cannot be started) by the tick of the time, either because its scheduling is non-constant or because its sampling interval is longer than 1 hour or because there is failure, its outputs are inferred, if possible, in an alternative way (often just reproducing the outputs of the previous hour) and its execution is re-planned for the following tick.

Once obtained, the data are arranged according to different criteria: separations are made between physical-chemical and microbiological features, and between quantitative and qualitative ones [10].

3.2.3 Physical and chemical features

Among the available physical and chemical features, the most relevant ones used by the DAI-DEPUR+ system are selected on the basis of human experience, tradition and utility measures. These features are not problematic and their modeling and application both in chemical engineering and artificial-intelligence systems are well documented in bibliography.

3.2.4 Microbiological features

The modeling of microbiological features exists in the scope of biological disciplines, but it has not yet been integrated into a decision support system dedicated to environmental issues, such as the DAI-DEPUR+ system. In this section we describe the methodology followed in the knowledge acquisition related to DAI-DEPUR+ [10].

In a WWTP, the identification of the microorganisms existing in the activated sludge is generally carried out in the laboratories of the plant and generates qualitative off-line data (e.g., presence of Paramecia species or diversity of Ciliate). Using an automatic quantitative analysis of digital images, for microorganism recognition and counting is a possibility for the future.

After the identification, a comparative study of microorganism communities of different treatment-plants is accomplished, to understand what can be the influence of biological variability at a geographical level. A set of microbiological features is then selected to be used by the system. For a high performance to be maintained throughout the domain (the different WWTPs), this feature set needs to be widespread enough to have a representational data-base with a relatively abundant number of instances. Referring to portability, the parameters available only in the minority of the treatment plants are not very useful in the development of the main knowledge-bases of the system, but they can be used as specific-domain knowledge by specially developed modules. Missing and incomplete information does not represent a problem in principle, but only a factor of increasing uncertainty.

3.3 Diagnosis layer

Once all data have been interpreted, the use of diagnostic knowledge-bases begins. Diagnosis is basic in the decision making of wastewater treatment. And the diagnosis layer is the one with most resources

allocated. The knowledge bases model the particular kind of WWTP from which the data are coming. The diagnosis is based on different reasoning models and the ontology.

3.3.1 Knowledge-based paradigm

In the DAI-DEPUR+ system there are a numerical control module and two AI knowledge-bases which:

- detect when the plant is in a *normal* state or in a *standard abnormal* state, such as bulking, storm or foaming states, and
- contribute to manage the general wastewater-treatment operation in these cases.

This routine management is carried out through automatic-control algorithms, case-based reasoning and rule-based reasoning. Case-based reasoning is often able to model also specific features and particular states of the treatment plant (*nonstandard abnormal* states), and to learn from past situations occurring in the treatment plant itself. This would account for the potential difference in individual treatment-plants due to deviations in parameters such as inflow, meteorology, neighboring industries and local life-style.

3.3.2 Ontology

An ontology is integrated with the KBSs mentioned in section above. With this ontology it is possible to capture, understand and describe the knowledge about the whole physical, chemical and microbiological environment of a WWTP.

The goal of the integration of the ontology is to create a model that:

1. provides a shared terminology for the wastewater domain that each agent can jointly understand and use; e.g., the shared term *descriptor* unifies the terms *variable*, *feature*, *attribute* and *parameter*, which are used by different agents to refer to the same concept;
2. defines the meaning of each term (part of the semantics) in a precise and as unambiguous manner as possible; e.g., 'the term *descriptor* refers to attributes which describe environmental conditions, such as the appearance of microorganism flocks or of the water surface of the clarifier';
3. encodes in an environmental decision-support system, for the first time, a deep microbiological knowledge; e.g., the taxonomy of the microorganisms which live in WWTPs;
4. links concepts with taxonomic / hierarchical relations; e.g., 'Nocardia *is a* Actinomycetes';
5. implements the semantics in a set of axioms that will enable the ontology to automatically deduce the answer to many questions about the wastewater domain, such as cause-effect questions; e.g., the axiom 'Actinomycetes *is cause of* Foaming sludge';
6. has axioms which permit diagnosis-impasse solving;
7. uses the Ontolingua environment for depicting concepts in a graphical context;
8. will integrate with some temporal reasoning, based on transition networks, to obtain a qualitative simulation of the evolution of WWTP states.

Axioms We approach goals 5 and 6 by defining a set of axioms (or rules) that describe wastewater processes. Axiom deductions should be determined by a set of questions used to decide the competence of the ontology's representation. Since there does not exist a standard for determining the competence of a model, we will define a set of questions about wastewater processes and the axioms used to answer them.

Basic entities The basic entities in the ontology are represented as objects with specific properties and relations. Objects are structured into a taxonomy and the definitions of objects, attributes and relations are specified according to the Ontolingua version of the frame ontology.

The hierarchical structure and the axioms of the ontology can help to diagnose the situation in case of *impasse* of the other KBSs.

The ontology is normally static. It activates its inference mechanisms (axioms) only under specific petitions from the diagnosis integrator (see next section). The result of the inference of the ontology is:

- an answer about the diagnosis impasse (e.g.: 'We have a foaming situation' or 'I do not have information to solve the impasse'),
- an explanation of the answer (e.g.: 'I received information related to the answer from the activation of the following axioms ...' or 'The answer was obtained searching the following classes ...').

The activation of the ontology always means that there was an impasse in KBS diagnosis. If the answer of the ontology to the petition is 'I do not have information to solve the impasse', then a primary alarm is activated.

3.3.3 Diagnosis integration

The rule-based expert system (RBES) and the case-based reasoning system (CBRS) work in parallel and they both produce as output a diagnosis on the state of the plant. This output is passed to the diagnosis integrator, a subsystem between the diagnosis and the decision support layers.

General integration schema If the diagnosis of the two KB systems is the same, it is passed to the decision support layer. If the diagnoses exist and are different, the system prioritizes as follow:

- If the case library contains a predefined minimum historical series and the case similarity is higher than a predefined value, the case-based reasoner's diagnosis prevails.
- Otherwise, the rule-based expert system's diagnosis prevails.

In case of impasse (no diagnosis), DAI-DEPUR+ turns first to the ontology and then, if it fails, to the plant manager, demanding an off-line diagnosis based on their microbiological deep knowledge. This external solution is learned.

Detailed integration schema

1. $\notin$ CBRS diagnosis and $\notin$ RBES diagnosis: impasse, the diagnosis integrator turns to the ontology.
2. $\notin$ CBRS diagnosis and $\in$ RBES diagnosis:
   - RBES diagnosis-certainty $\geq a$: RBES diagnosis passed to decision support layer.
   - RBES diagnosis-certainty $< a$: impasse, the diagnosis integrator turns to the ontology.
3. $\in$ CBRS diagnosis and $\notin$ RBES diagnosis:
   - CBRS case-similarity $\geq b$: CBRS diagnosis passed to decision support layer.

- CBRS case-similarity $< b$: impasse, the diagnosis integrator turns to the ontology.

4. $\in$ CBRS diagnosis and $\in$ RBES diagnosis:
   - CBRS case-similarity $\geq b$: CBRS diagnosis passed to decision support layer.
   - CBRS case-similarity $< b$:
     - RBES diagnosis-certainty $\geq a$: RBES diagnosis passed to decision support layer.
     - RBES diagnosis-certainty $< a$: impasse, the diagnosis integrator turns to the ontology.

Example (case 1) We have a certain perception state *A*. The case-based reasoning system (CBRS) and the rule-based expert system (RBES) activate.

The CBRS finds a case similar to state *A*. The similarity value is 0.1 and it is less than the minimum acceptable value *b* (e.g. b=0.2). Therefore there is no diagnosis output from the CBRS.

The RBES finds no rules leading to a diagnosis starting from state *A*. Therefore there is no diagnosis output from the RBES.

The diagnosis integrator acknowledges a case of missing diagnosis from the KBSs and send a petition to the ontology with the description of state *A*.

The output of the ontology is: Answer = 'We have a foaming situation', Explanation = 'I received information related to the answer from the activation of the following relation (Nocardia *is a* Actinomycetes) and the following axioms (Actinomycetes *is cause of* Foaming sludge)'.

Comment: State *A* is characterized by a strong presence of Nocardia bacterium, but this bacterium has no direct relation (in the static knowledge-base of the RBES) to any state of the WWTP. Even in the ontology the Nocardia class has no link to any state of the WWTP, but its parent class (Actinomycetes) has a cause-effect link to the general state of the WWTP 'Foaming sludge'. One reason for this could be that Nocardia is not causing foaming, but another (not detected) bacterium of the same taxonomic class is.

The diagnosis integrator receives the diagnosis from the ontology and pass it to the decision support layer.

3.4 Decision support layer

Eventually, we describe the supervisory level of the DAI-DEPUR+ environmental decision-support system. Once the diagnoses of the reasoners (case-based and rule-based) and possibly of the ontology for the management of the wastewater treatment have been integrated, it selects an actuation.

Real time This layer runs always in real time and it is very robust in this sense, simply because of the fact that the entire system's minimum-cycle time is very long with respect to the calculations done by the decision support layer.

3.4.1 Prediction

The result of diagnosis integration (carried out among case-based reasoning, rule-based reasoning and the ontology) serves as input for the prediction phase. A subsystem, based on transition networks, predicts various alternative evolutions of the state of the WWTP. A second subsystem evaluates these alternatives. The result is passed on to the actuation selector.

Actuation selection Actions to be carried out are selected. Often, action schemas are already included in the diagnosis result.

4 CONCLUSIONS AND FUTURE WORK

In this paper we presented the DAI-DEPUR+ decision support system, integrating a rule-based expert system, a case-based reasoner and an ontological knowledge-base. This system is able to model the information about a wastewater treatment process. The main improvements of DAI-DEPUR+ system with respect to existent system are:

- Impasse situations in existent systems are solved by the ontology.
- While in existent systems there is no modeling of wastewater microbiology, in this new system the microbiological component is modeled by the ontology.
- DAI-DEPUR+ presents a novel integration between KBSs and ontologies in a real world application.
- DAI-DEPUR+ facilitates its own portability.
- DAI-DEPUR+ incorporates cause-effect reasoning.
- DAI-DEPUR+ will incorporate predictive skills.
- It will be possible a semi-automatic generation of a static KB.

Acknowledgments

This research is supported by the Generalitat de Catalunya (grant 1997FI 00630 UPC PG).

REFERENCES

[1] Alamán, X., Romero, S., Aguirre, C., Serrahima, P., Muñoz, R., López, V., Dorronsoro, J. and de Pablo, E., 1992. *MIP: A Real Time Expert System*, in the Proceedings of the 8th Conf. On Artificial Intelligence Application (CAIA-92), Monterrey, California.

[2] Avouris, N.M., 1995. *Co-operating Knowledge-Based Systems for Environmental Decision-Support*, in Knowledge-Based Systems 8(1) (1995), pp. 39-53.

[3] Baeza, J., Gabriel, D. and Lafuente, J., 1998. *An Expert Supervisory System for a Pilot WWTP*, in ECAI 98 - W7 (BESAI98) workshop notes, pp. 25-37.

[4] Barnett, M.W. and Andrews, J.F., 1990. *Knowledge Based Systems for Operation of Wastewater Treatment Processes*, in Instrumentation, Control and Automation of Water and Wastewater Treatment and Transport Systems, Proceedings of the 5th IAWPRC workshop, Yokohama and Kyoto, pp. 211-18, Pergamon Press, New York 1990.

[5] Beccari, M., 1991. *Principi dei trattamenti biologici*, in Canziani, R. (editor) *Trattamento delle acque di rifiuto*, pp. 99-113, Istituto per l'Ambiente, Milano, Italy 1991.

[6] Borst, P., Akkermans, H. and Top, J., 1997. *Engineering Ontologies*, in International Journal of Human-Computer Studies 46, pp. 365-406, 1997.

[7] Branting, K., Hastings, J.D. and Lockwood, J.A., 1997. *CARMA: a pest management advisory system*, in AI Applications 11(1) (1997), pp. 29-48.

[8] Brooks, R.A., 1986. *A robust layered control system for a mobile robot*, in IEEE J. Rob. Aut. (RA) 2(1) (1986), pp. 14-23.

[9] Buchanan, B. and Smith, R., 1988. *Fundamentals of expert systems*, in Annual Reviews on Computer Science 3 (1988), pp. 23-58.

[10] Comas, Q., R-Roda, I., Ceccaroni, L. and Sànchez-Marrè, M., 1999. *Semi-automatic learning with quantitative and qualitative features*, in Proceedings of the VIII Conference of the Spanish Association for Artificial Intelligence (CAEPIA-99), November 16-19, 1999, Murcia, Spain.

[11] Cortés, U. and Sànchez-Marrè, M. (editors), 1998. *Binding environmental sciences and artificial intelligence*, ECAI 98 - W7 (BESAI98) workshop notes.

[12] Damiani, A., 1991. *Caratteristiche dei liquami urbani*, in Canziani, R. (editor) *Trattamento delle acque di rifiuto*, pp. 11-42, Istituto per l'Ambiente, Milano, Italy 1991.

[13] Droste, R. L., 1997. *Theory and practice of water and wastewater treatment*, Wiley 1997.
[14] Dym, C.L. and Levitt, R.E., 1991. *Knowledge-based Systems in Engineering*, McGraw-Hill 1991.
[15] Farquhar, A., Fikes, R., Pratt, W. and Rice, J., 1995. *Collaborative Ontology Construction for Information Integration*, technical report KSL-95-63, August 1995, Knowledge Systems Laboratory, Department of Computer Science, Stanford University, Stanford, US.
[16] Farquhar, A., Fikes, R. and Rice, J., 1996. *The Ontolingua server: a tool for collaborative ontology construction*, technical report KSL-96-26, Knowledge Systems Laboratory, Stanford University, Stanford, US.
[17] Fedra, K., 1993. *Expert systems in water resources simulation and optimization*, in Marco, J.B. et al. (editors) *Stochastic hydrology and its use in water resources systems simulation and optimization*, pp. 397-412, Kluwer Academic Publishers, The Netherlands 1993.
[18] Fedra, K., 1995. *Decision support for natural resources management: models, GIS and expert systems*, in AI Applications 9(3) (1995), pp. 3-19.
[19] Fikes, R., Farquhar, A. and Rice, J., 1997. *Tools for Assembling Modular Ontologies in Ontolingua*, technical report KSL-97-03, Knowledge Systems Laboratory, Stanford University, Stanford, US.
[20] Finch, F.E., Oyeleye, O.O. and Kramer, M.A., 1990. *A Robust Event-Oriented Methodology for Diagnosis of Dynamic Process Systems*, in Computers & Chemical Engineering 14(12) (1990), pp. 1379-96.
[21] Gómez-Pérez, A., 1998. *Knowledge sharing and reuse*, in the Handbook of Applied Expert Systems, pp. 10.1-10.36, CRC Press 1998.
[22] Gómez-Pérez, A., Fernández, M. and de Vicente, A., 1996. *Towards a method to conceptualize domain ontologies*, in the proceedings of the workshop on Ontological Engineering, ECAI'96, Budapest, Hungary, pp. 41-52.
[23] González and Dankel, 1994. *The engineering of knowledge-based systems*, Prentice-Hall 1994.
[24] Gruber, T.R., 1993. *A Translation Approach to Portable Ontologies*, in Knowledge Acquisition 5(2), pp. 199-220, 1993.
[25] Guariso, G. and Werthner, H. (editors), 1989. *Environmental Decision Support Systems*, Ellis Horwood - Wiley 1989.
[26] Jackson, P., 1990. *Introduction to expert systems*, Addison Wesley 1990.
[27] Jones, D.M., Bench-Capon, T.J.M. and Visser, P.R.S., 1998. *Methodologies for Ontology Development*, IT&KNOWS - Information Technology and Knowledge Systems, 15th IFIP World Computer Congress, Vienna (Austria) and Budapest (Bulgaria).
[28] Kaelbling, L.P., 1990. *An architecture for intelligent reactive systems*, in Allen, J., Hendler, J. and Tate, A. (editors) *Readings in Planning*, pp. 713-28, Morgan Kaufmann 1990.
[29] Krovvidy, S. and Wee, W.G., 1991. *Wastewater treatment systems from case-based reasoning*, in Machine Learning 10 (1991), pp. 341-63.
[30] Mason, C. (editor), 1995. *Workshop on Artificial Intelligence and the Environment*, in IJCAI-95 Workshop Program Working Notes.
[31] Matwin, S., Charlebois, D., Goodenough, D.G. and Bhogal, P., 1995. *Machine learning and planning for data management in forestry*, in IEEE Expert Systems 10 (5) (1995).
[32] Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T. and Swartout, W.R., 1991. *Enabling Technology for Knowledge Sharing*, in AI Magazine (winter 1991), pp. 36-56.
[33] Novotny, V., Jones, H., Feng, X. and Capodaglio, A.G., 1990. *Time Series Analysis Models of Activated Sludge Plants*, in Water Science & Technology 23(4-6) (1990), pp. 1107-16.
[34] Ortolano, L., LeCoeur, G. and MacGilchrist, R., 1990. *Expert System for Sewer Network Maintenance: Validation Issues*, in Journal of Computing in Civil Engineering 4(1) (1990), pp. 37-54.
[35] Patry, G. and Chapman, D. (editors), 1989. *Dynamic modelling and expert systems in wastewater engineering*, Lewis Publisher, Chelsea, MI, USA 1989.
[36] Sànchez-Marrè, M., 1995. *DAI-DEPUR: an integrated supervisory multi-level architecture for wastewater treatment plants*, Ph.D. Thesis, Software Department, UPC, Barcelona, Spain.
[37] Sànchez-Marrè, M., Cortés, U., R-Roda, I. and Poch, M., 1998. *Case Learning through a Similarity Measure for continuous domains*, in ECAI 98 - W7 (BESAI98) workshop notes, pp. 39-53.
[38] Sanz, R., Aguilar, J., Sierra, C., Godó, L. and Ollero, A., 1989. *Adaptive control with a supervisor level using a rule-based inference system with approximate reasoning*, in Artificial Intelligence in Scientific Computation: towards 2nd generation systems, IMACS, 1989.
[39] Serra, P., 1993. *Development of a knowledge-based system for control and supervision of urban wastewater treatment plants*, Ph.D. Thesis, Chemistry Department, UAB, Barcelona, Spain.
[40] Shepherd, A. and Ortolano, L., 1996. *Water-supply system operations: Critiquing expert-system approach*, in J. Of Water Resources Planning and Management 122(5) (1996), pp. 348-55.
[41] Stephanopoulos, G., 1990. *Artificial intelligence in process engineering: current state and future trends*, in Computers & Chemical Engineering 14 (1990), pp. 1259-70.
[42] Wright, J., Wiggins, L., Jain, R. and John Kim, T. (editors), 1993. *Expert systems in Environmental Planning*, Springer-Verlag, Berlin-Heidelberg-NewYork 1993.
[43] Yang, C.T. and Kao, J.J., 1996. *An expert-system for selecting and sequencing wastewater treatment processes*, in Water Science & Technology 34(3-4) (1996), pp. 347-53.
[44] Zhu, X.X. and Simpson, A.R., 1996. *Expert System for Water Treatment Plant Operation*, in Journal of Environmental Engineering 122(9) (1996), pp. 822-29.