Available online at www.sciencedirect.com
SCIENCE DIRECT®

Environmental Modelling & Software

Environmental Modelling & Software XX (2004) XXX–XXX

www.elsevier.com/locate/envsoft

OntoWEDSS: augmenting environmental decision-support systems with ontologies

Luigi Ceccaroni [a,b,*], Ulises Cortés [b], Miquel Sànchez-Marrè [b]

[a] *Fujitsu Laboratories of America, 595 Lawrence Expressway, Sunnyvale, CA 94085, USA*
[b] *Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya Mòdul C5, Campus Nord, C/Jordi Girona 1-3, 08034 Barcelona, Catalunya, Spain*

Received 15 October 2002; received in revised form 27 January 2003; accepted 5 March 2003

Abstract

This paper characterizes part of an interdisciplinary research effort on AI techniques applied to environmental decision-support systems. The architectural design of the OntoWEDSS decision-support system for wastewater management is presented. This system augments classic rule-based reasoning and case-based reasoning with a domain ontology, which provides a more flexible management capability to OntoWEDSS. The construction of the decision-support system is based on a specific case study. But the system is also of general interest, given that its ontology-underpinned architecture can be applied to any wastewater treatment plant and, at an appropriate level of abstraction, to other environmental domains. The OntoWEDSS system helps improve the diagnosis of faulty states of a treatment plant, provides support for complex problem-solving and facilitates knowledge modeling and reuse. In particular, the following issues are dealt with: (1) modeling information about wastewater treatment processes, (2) clarifying part of the existing terminological confusion in the domain, (3) incorporating ontology-modeled microbiologic knowledge related to the treatment process into the reasoning process and (4) creating a *decision-support system* that combines information through a novel integration between knowledge-based systems and ontologies.

Keywords: Ontologies; Environmental modeling; Artificial intelligence; Decision-support systems

1. Introduction

The goal of this research is to identify how the introduction of an advanced knowledge-representation system, an ontology (see Section 1.3), improves decision-support systems (DSSs) that employ rule-based reasoning (RBR) and case-based reasoning (CBR) approaches. Through the analysis of a specific DSS, however, we will discuss a general and controversial concept: ontologies used in reasoning. We present a system architecture which represents a new, interdisciplinary approach to the management of knowledge in a problem-solving process. Even if the application studied is specific, it can serve as a basis for any environmental system.

Most would agree that it is important to have systems that understand ontology basic concepts (such as *subclass* and *inverse*), and even better if we could state any logical principle and permit a system to reason (by interface) using these principles (Swartz and Hendler, 2001). For example: we know that, if in a *wastewater treatment plant* (WWTP) there is excessive proliferation of filamentous-bacteria, then we have a bulking situation. Let us define, for the sake of this example, “filamentous-bacteria excessive proliferation” as “*sludge-volumetric index* (SVI) greater than 140” or “total filamentous-bacteria greater than 90 m/ml”. A *smart* system can now follow this rule to make a simple deduction: “On September 24, 2000, SVI was 215, therefore we had bulking”. Another example: WWTP records show that on October 25, 1999, the concentration of *Microthrix parvicella* was 80 m/ml and the one of *Gordona* was 13 m/ml. The ontology states that *Microthrix parvicella* and

* Corresponding author. Knowledge Engineering and Machine Learning Group (KEMLG), Artificial Intelligence Section, Technical University of Catalonia, Campus Nord-Edifici C5, Jordi Girona 1-3, Barcelona, Catalonia, Spain. Tel.: +34-93-4017994; fax: +34-93-4017014.
E-mail address: luigi.ceccaroni@lycos.com (L. Ceccaroni).

1364-8152/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2003.03.006

Gordona are both different filamentous-bacteria. The built-in math rules state that 80 + 13 = 93 and that 93 is more than 90. And, as we know, if total filamentous-bacteria are greater than 90, we have bulking. The system puts all these logical rules together into a proof that “On October 25, 1999 we had bulking.” Once we begin to build systems that follow logic, we can use them to provide proofs. Scientists at some WWTPs could write logic statements; then, if they use the same ontologies and the premises of their statements are satisfied, other WWTPs could follow these statements (i.e. semantic links or decision trees, see Section 2.2) to construct proofs. While it is very difficult to create these proofs (it may require following tens or perhaps hundreds of links and rules), it is relatively easy to check them. In this way, we begin to build a network of information processors. Some of them merely provide data for others to use. Others are smarter and can use these data to build rules. The smartest processors are *heuristic engines* that follow all these rules and statements to draw conclusions (and can modify themselves in response to user input), then place their results back on the network (Swartz and Hendler, 2001). In summary, the defining feature of the use of ontologies in reasoning will be the ease with which PDAs, laptops, WWTPs desktops, servers, sensors and actuators will communicate with each other and with other machines in different WWTPs. This will facilitate the automation of decisions and actuations that previously had to be laboriously hand-processed.

This vision is partially achieved by prototyping a new DSS architecture, ontology-based wastewater environmental decision-support system (OntoWEDSS), and applying it to the domain of wastewater treatment. We show and explain how the addition of ontologies and semantics results in a more reliable and practical management of complex environmental problems. However, the performance of OntoWEDSS could not be evaluated against other similar systems (for the reasons stated in Section 5).

This paper is structured as follows. The remainder of this section provides general background on *environmental decision-support systems* (EDSSs), on previously developed DSSs in the domain of interest (wastewater management), on the difficulties related to these DSSs, and on ontologies. Section 2 illustrates how OntoWEDSS has been designed, depicts its layered architecture and explains the functioning of its components. Section 3 focuses on the role of ontologies in reasoning, while Section 4 provides information on how the system works. Section 5, discusses the performance of the system and, finally, Section 6 presents our conclusions.

1.1. Environmental decision-support systems

EDSSs are useful when dealing with complex environmental processes that are not easily modeled because the knowledge is still incomplete and uncertain. *Rule-based expert systems* (RBES), *case-based reasoning systems* (CBRS) and ontologies (see Section 1.3) have each proven to be able to cope, individually, with some known difficulties and to successfully face several problems related to the wastewater domain (Ceccaroni, 2001), and ontologies can also be used to improve knowledge-based system (KBS) development (van Heijst et al., 1997). The synthesis of these different modeling and reasoning systems can result in great improvements in decision-support (Cortés et al., 2001). The introduction of an ontological component in an EDSS, in particular, allows: (1) a more stable wastewater treatment operation through ontology-based supervision and (2) the *portability* of the management system of a WWTP.

WWTPs serve to decontaminate wastewaters prior to their discharge into a natural body of water. For that, they use techniques of physical, chemical and biological treatment. The wastewater treatment process is very *complex* (mainly because of the presence of microbial communities in the activated-sludge) and it is difficult to develop a reliable *supervisory technology* based only on a traditional chemical-engineering control approach. What is possible to develop is a *model* of the process. Different models can be found in the bibliography, with distinct levels of complexity in the characterization of the activated-sludge process; these models contribute to a better understanding of the process and are a powerful support tool for operators and plant managers to evaluate various scenarios by experimenting with simulations (Comas, 2000).

1.2. Previous prototype DSSs for wastewater management and their problems

An initial prototype for supervision and management, called DAI-DEPUR, was developed by Sànchez-Marrè et al. (1996). Its architecture combines several reasoning techniques in a single framework, such as learning, knowledge acquisition, distributed problem-solving, RBR and CBR. In DAI-DEPUR, four categories are distinguished from the domain model point of view: data, knowledge, situations and plans. From the management task point of view, five categories are considered: evaluation, diagnosis, supervision, actuation and learning. DAI-DEPUR is implemented in LISP and most of its subsystems can be executed in parallel. The authors claim that a *supervisory system,* through predetermined plans and actuations, is more efficient than other kinds of distributed AI systems, such as *blackboard systems* and *contract nets,* to deal with WWTP’s *abnormal* situations, such as *storm*, *bulking* and *toxic load.* Another prototype for supervision is BIOMASS (Comas, 2000), a system applied to WWTPs that integrates the reasoning capabilities of an RBES and a CBRS within a DSS that manages data and decision-making. The core and main modules of the system have been developed in the proprietary Gensym's G2 object-oriented shell, which makes difficult further independent research-efforts by others, due to unjustifiable high costs. On the other hand, BIOMASS is the only system that has actually been tested as a real application in the field and has been installed and run for a substantial time on a working, full-scale WWTP. G2 is a user-friendly development environment, which embodies its own inference engine and controls on-line and off-line data acquisition, database management, RBR and CBR. Apart from this, BIOMASS includes *supervisory control and data acquisition* (SCADA) and *programmable logic circuits* (PLC) networks with basic control algorithms. Knowledge representation is achieved through classification and rule application. BIOMASS has an object base that is hierarchically structured with classes and subclasses. The scheme of a specific WWTP is built from the instantiations of these classes connected among each other. Finally, the DAI-DEPUR+ system (Ceccaroni, 2000) needs to be cited. It derives from the DAI-DEPUR system, is its direct evolution through the addition of an ontology, and is the precursor of the system presented in this paper, of which it represents a non-refined, early version.

All the systems described above are at an experimental stage of development; they still have problems and can be improved. In Section 5.1, where necessary concepts will be introduced, we explain in detail how OntoWEDSS improves on these systems. Here, we simply point out their main difficulties:

- They have multiple diagnosis modules, but these modules are not integrated.
- Confusion about terminology is likely to happen. A semantic integration of information does not exist: a problem that goes beyond syntactic integration (dealt with by BIOMASS and in part by DAI-DEPUR) is the mapping of semantics of terms from different information sources (such as different WWTPs), even when these terms have been expressed using the same syntactic structures.
- In DAI-DEPUR, there is no modeling of wastewater microbiology.

1.3. Ontologies

In the introduction of the paper, we anticipated how ontologies may affect and improve decision-support systems. In this section, we explain more formally what ontologies are and what problems can arise from knowledge sharing.

Ontologies are being developed to facilitate knowledge sharing and reuse. With respect to the research involved in this study, ontologies can provide: (1) a shared and common understanding of the knowledge domain that can be communicated among agents and application systems, and (2) an explicit conceptualization that describes the semantics of the data (Fensel et al., 2000). A recent comprehensive document covering the main aspects of ontologies in AI research is the technical roadmap of the ontology field in Europe and worldwide produced by the OntoWeb project (OntoWeb project, 2002). With respect to EDSSs and WWTPs, we would like to highlight two characteristics of ontologies:

- On one hand, ontologies represent the first step on the way to *real portability* of a system towards other similar domains: they could be effectively employed to address the problem of general model-construction (*generalization*).
- On the other hand, it is possible to instantiate/adapt an ontology to the specific configuration of a WWTP and to automatically construct and validate specific models (*specification*).

Some could note that when we start sharing and reusing external ontologies and logic statements, security problems arise. We do not deal with such problems in the paper, but we briefly explain here how it might be done. Who, in fact, would trust a system using such external statements? These security and trust problems are in general dealt with by digital signatures (W3C, 1998). Based on the work in mathematics and cryptography, digital signatures provide proof that a certain person wrote (or agrees with) a document or statement. Therefore, all ontologies need to be digitally signed. That way, we can be sure about who wrote them. We simply tell our system whose signatures to trust and whose not to. Each WWTP can set its own levels of trust and can decide how much of what it reads to believe. A similar concept, parallel to the one of trust, is similarity. As an example, if we know that another WWTP is a twin plant with respect to ours, we associate a high degree of confidence to the statements coming from that plant. So, first we check the digital signature and then the similarity. Now, it is highly unlikely that we will trust enough people to make use of most of the available ontologies. That is where the *Web of Trust* comes in. We tell our system that we trust another WWTP. That WWTP trusts other WWTPs, and so on. As these trust relationships fan out from us, they form a *Web of Trust*. And each of these relationships has a degree of trust (or distrust) associated with it (Swartz and Hendler, 2001). The system takes all these factors into account when deciding how trustworthy a piece of information is. Information with different degrees of trust can be used at different levels of reasoning. One thing is to just detect bulking, a totally different one is to find the sequence of actuations to solve a bulking problem, together with a complete explanation and an assessment of the reliability of the factors involved in the decision.

2. OntoWEDSS

OntoWEDSS is a research tool built to explore the possibilities and the potential of introducing ontologies into decision-support systems, using an environmental domain as a case study. In this section, we describe the OntoWEDSS system and its different functional units, while in Section 3 we will focus on the specific role of ontologies.

The architecture of OntoWEDSS integrates various kinds of data and several AI techniques. Given an adequate amount and type of data, it is flexible enough to deal with the complexity of the wastewater treatment process. In OntoWEDSS, the domain is represented in detail and the evolution of micro-organism communities (a key element in the biological treatment process) is taken into account. With OntoWEDSS, it is possible to capture, understand and describe the knowledge on the whole physical, chemical and microbiological environment of a WWTP. OntoWEDSS incorporates wastewater microbiological knowledge into the reasoning process and represents cause-effect relations. It makes use of an ontology for domain modeling and for the clarification of the existing terminological confusion in the wastewater domain (see Fig. 1). OntoWEDSS discovers problematic states in the wastewater treatment domain and uses different reasoning systems (rule-based, case-based and ontology-based), which interoperate among themselves. Finally, OntoWEDSS resolves existing *reasoning-impasses*, such as lack of diagnosis.

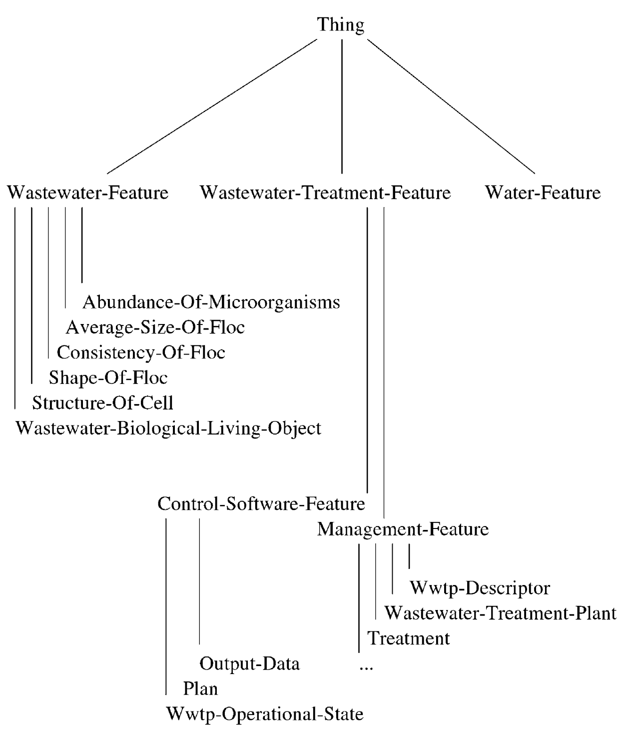

Fig. 1. Top-level categories in the October 2001 version of the WaWO ontology. The full classification is available at the Ontolingua ontology server (http://www.ksl-svc.stanford.edu).

The *input* (from the user or a file) for modeling and execution in OntoWEDSS is represented by (1) the list of descriptors to use and (2) the descriptor values of a new problem. The user can take advantage of a predefined set of descriptors and can define new ones (in this second case, however, only case-based reasoning is readily available for diagnosis). Optionally, the weight of the descriptors can be provided. What follows is an *example* of input descriptors[1] together with their range of possible values: *foam presence at aeration tank* (none, very little, some, abundant, very abundant) *floc appearance at clarifier* (deflocculation, floating sludge, denitrification), *water qualitative-assessment at aeration tank* (bad, good, very good). The *output* (to the user) of OntoWEDSS execution is represented by (1) a diagnosis of the current state of the WWTP (with a reliability factor), (2) a trace of the reasoning carried out, and (3) a list of actions to be taken according to the situation diagnosed.

The architecture of the system (see Fig. 2) includes many distinct, specialized subsystems (such as RBR, CBR, ontology, planning and actuation), which are defined and then grouped to accomplish three conceptual tasks, whose details are as follows: *perception* (data gathering), *modeling and diagnosis* (including learning), *decision-support* (prediction, evaluation of alternative scenarios, advising, actuation and supervision).

2.1. Perception

The OntoWEDSS system's domain physically consists of a wastewater treatment plant. In particular, all the physical, chemical and biological measurements for this research were gathered in treatment plants located in Spain. The time-scales of the treatment processes are long, so that the perception and the supervision decisions easily fit between sampling cycles. The perception task includes the process of information integration. Two integration levels have to be dealt with in order to achieve a completely integrated access to information (Stuckenschmidt, 2000). The first level is the *structural integration*, which is concerned with network technology and communication protocols, ensuring that the different information sources can physically communicate. Once the sources can physically exchange information, they must agree on a common syntax for exchanging such information (*syntactic integration*). An example of the

[1] OntoWEDSS uses dozens of physical, chemical and biological descriptors (see also Table 1).

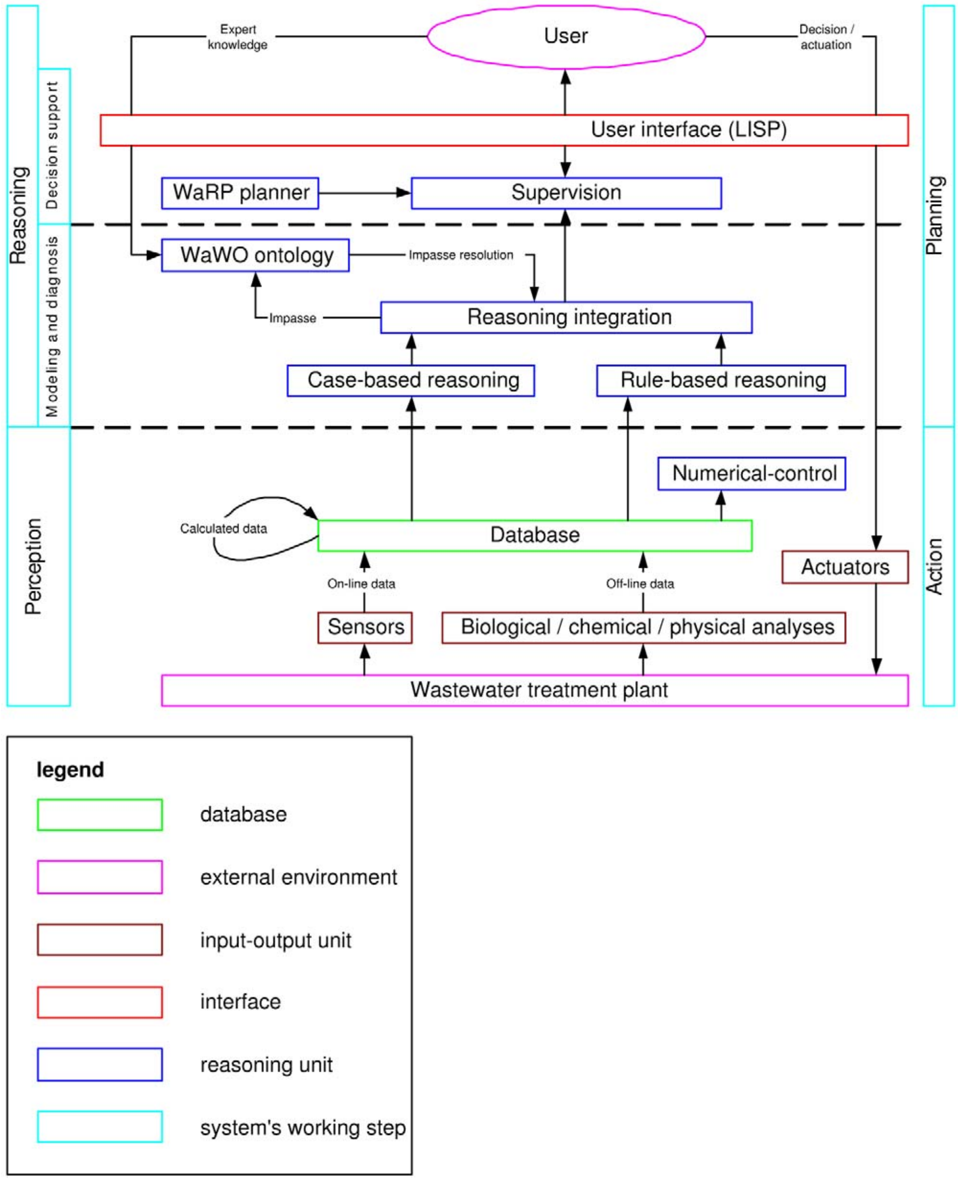

Fig. 2. The OntoWEDSS architecture.

list-based *syntax* of domain definition in OntoWEDSS is as follows.

```
("Wastewater treatment plant of Granollers" 21
 ...
 ("TSS-I" "quantitative" 0.7 3
   "0" "230" "Low"
   "231" "330" "Normal"
   "331" "750" "High")
 ("SVI-AT" "quantitative" 0.8 3
   "0" "50" "Low"
   "51" "100" "Normal"
   "101" "1000" "High")
 ("Filam-Dominant-AT" "qualitative" 1 0.8 20
   "Haliscomenobacter-Hydrosis"
   "Microthrix-parvicella"
   "Gordona"
   ...)
 ("Diagnosis" "qualitative" 1 1 35
   "Deflocculation"
   "Electrical-Blackout"
   .........
   "Bulking-Sludge-Filamentous"))
```

This is the domain structure used for input. Some of the descriptors, such as *Diagnosis* and the ones referring to actuation (not shown), are used only when training the CBRS and are empty when introducing a new problem. Input is currently untyped and can be transformed into strongly typed, if needed, by not allowing the augmentation of WaWO's classes. If communication among different WWTPs is involved a third level of *semantic integration* is necessary, as explained in Section 1.

Data are distinguished into three types: on-line data, off-line data and calculated data (see Table 1):

1. *On-line data* are always quantitative. On-line data acquisition will be directly accomplished by the PLC network of the WWTP.
2. *Off-line data* can be quantitative and qualitative. Off-line data, in general, are read by OntoWEDSS as an ASCII file. There are, though, several microbiological descriptors, whose values are determined only once a week. These data are introduced into the OntoWEDSS by the user through the API of the RBES module.
3. *Calculated data* are a combination of quantitative data which allow the assessment of global-process descriptors, such as Sludge-Residence Time, Sludge-Volumetric-Index, Food-To-Micro-Organism-Ratio. With respect to micro-organism information, the relative abundance of microbiological species is calculated (assigning to each species-descriptor one of the following ordered values: none, few, some, equilibrium, abundant, excessive).

Before being used for reasoning, the data require a number of processing procedures (not discussed here) to be validated, integrated into a uniform time-scale and discretized.

2.2. Modeling and diagnosis

The OntoWEDSS system uses its internal knowledge bases and inference mechanisms to process information about a WWTP. It diagnoses the ongoing state of the treatment plant and predicts the evolution of that state. The diagnosis is based on a model of the processes happening in the WWTP. This phase is characterized by a chicken-and-egg paradox: the situations (set of descriptors' values) cannot be defined without first knowing what diagnostics they correspond to; and most diagnostics can be hard to define as such until the corresponding situations have been identified. To overcome this problem, experts often have to use trial-and-error methods. The situations are represented under the form of *decision trees* (Comas, 2000). In fact, all symptoms, facts, procedures, relations and most logic statements and semantic links used in diagnosis and decision-support can be graphically represented with decision trees (see Fig. 3). They represent expert's procedural knowledge and decision-making behavior. These trees correspond to causal paths of interactions from symptoms to problems, using nodes interconnected by arcs. Each node refers to a descriptor or a test about a descriptor, whereas each arc corresponds to a possible value for that descriptor or

Table 1
Descriptors of WWTPs as they appear in the WaWO ontology (AT = aeration tank; S = settler; P = primary - treatment's effluent; I = inflow; E = WWTP's effluent)

Type	Descriptor	Sampling location
Off-line		
Qualitative descriptors		
	Appearance-Floc	AT and S
	Foam-presence	AT and S
	Water-odor	AT and S
	Water-quality	AT and S
	Biodiversity-of-ciliates	AT
	Biodiversity-of-filamentous-bacteria	AT
	Biodiversity-of-microfauna	AT
	Dominant-filamentous-bacteria	AT
	Flocs-morphology	AT
	Overall-evaluation-of-floc-quality	AT
	Microfauna-amoebae	AT
	Microfauna-ciliates	AT
	Microfauna-filamentous-bacteria	AT
	Microfauna-flagellates	AT
	Microfauna-metazoa	AT
	Microfauna-unidentified-ciliates	AT
	Total-filaments	AT
Quantitative descriptors		
	Ammonia	I, P and E
	BOD	I, P and E
	COD	I, P and E
	Chlorine	I, P and E
	Conductivity	I, P and E
	Greases	I, P and E
	Inhibitors	I, P and E
	Metals	I, P and E
	N-Total (TN)	I, P and E
	Nitrate	I, P and E
	Nitrite	I, P and E
	Oils	I, P and E
	Phosphate	I, P and E
	Phosphorus	I, P and E
	Temperature	I, P and E
	Total-Kjeldahl-Nitrogen (TKN)	I, P and E
	Total-suspended-solids (TSS)	I, P and E
	Turbidity	I, P and E
	Mixed-liquor-SS	AT
	Mixed-liquor-volatile-SS	AT
On-line		
	Dissolved-oxygen	AT
	pH	I and E
	Sludge-flow-rate	I, P, E, AT
	Water-flow-rate	I, P, E, AT
Calculated		
	%-BOD-removal	–
	%-COD-removal	–
	%-TSS-removal	–
	Food-to-micro-organism-ratio	–
	Hydraulic-residence-time	–
	Sludge-residence-time	–
	Sludge-volumetric-index (SVI)	–
	Relative abundance of microbiological species	–
	Predominant filamentous	–
	Predominant protozoan	–

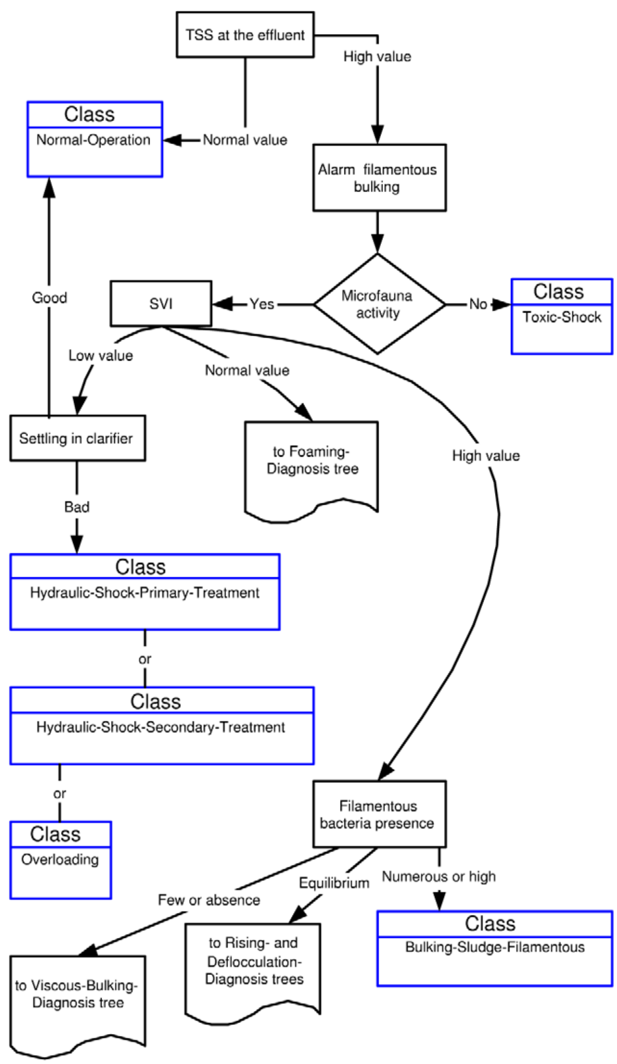

Fig. 3. Main decision tree for filamentous-bulking problems (simplified).

that test. There are different kinds of decision trees and here we refer only to the ones used for diagnosis. In a diagnosis decision-tree, leaf nodes represent a subclass of the WaWO's class *WWTP-Operational-State* (see Ceccaroni, 2001: p. 119). The translation of the knowledge contained in decision trees into *rules* for the RBES is direct. For example, the arc and two nodes at the top of Fig. 3 identify the rule *IF TSS at the effluent has a high value THEN an alarm for filamentous bulking should be activated*. The decision trees we use are not active objects: the possibility to change cuts between modalities of values exists (for example, we can decide that the cut between *high value* and *normal value* for TSS is 10 instead of 20 mg/l) but, for the trees to be reused in other WWTPs, a practical, automatic way to change the descriptors in the nodes (or to change the destination of the arcs) is needed. Otherwise, the rule system reflecting the decision trees is too static for adaptation.

Once the process is modeled, three modules (*RBR*, *CBR* and the *ontology*) are used for *diagnosis*. For this task, the RBES and the CBRS work independently and they both produce as output a diagnostics about the state of the plant. This output is passed to a diagnosis integrator, which shows the two outputs (together with an associated confidence value) to the user and then starts the reasoning schema illustrated in Fig. 4. If the diagnostics of the two KBSs is the same, this result is used for the decision-support task. If the diagnostics exist and are different, the system prioritizes as follows. If the case similarity is higher than a predefined threshold *b*, the case-based reasoner's diagnostics prevails. Otherwise, the RBES's diagnostics prevails. In case of impasse (no diagnosis), OntoWEDSS turns first to the ontology and then, if it fails, to the plant manager, demanding a diagnosis based on their microbiological deep knowledge. This external solution is used in the learning process of the CBRS.

The *modeling and diagnosis* task is accomplished through programs implemented in Allegro Common Lisp 5.0.1, a programming environment developed by Franz, Inc., which lets generate an application with a graphical interface. The three modules used for this task are now briefly described.

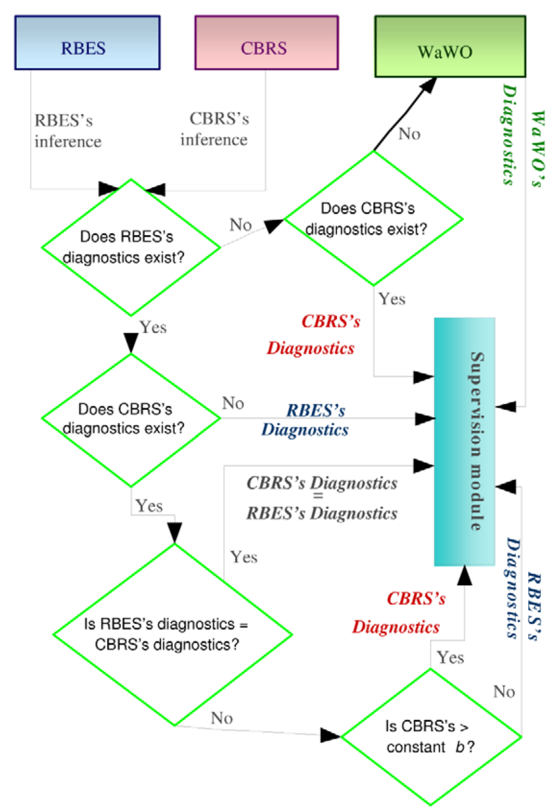

Fig. 4. Data flow of general diagnosis-integration.

2.2.1. RBES

This module is a shell, enabling the development of an expert system based on rules. The rule system is designed to be implemented in *two separate layers*: a more general one, which can be reused across WWTPs, and a more specific one to be used only in a particular WWTP. The user defines the set of data/facts (data base) and the set of rules (knowledge base). The diagnostics comes from the execution of the RBES with the data and rules introduced. An interaction with the user during execution is possible. The result is presented together with an explanation which shows the reasoning sequence. The elements which compose this module are the same as a typical RBES: database, knowledge base (KB) formed by rules and meta-rules, inference engine, user interface with question–answer management, explanation module which traces the reasoning of the system, and knowledge engineer interface which connects the engineer with the KB and the database. Here, we do not characterize each element of this module nor describe the specific implementation of uncertainty in the RBES.

2.2.2. CBRS

This module is a shell for the definition of a case-based reasoning system. To represent domain knowledge, *descriptor–value pairs* are used. To represent a problem, a set of descriptors, with their associated values, is used. These descriptors are a subset of WaWO's vocabulary. Depending on the particular WWTP, a different subset can be chosen. The structure of a descriptor is as follows (see Fig. 5): name, weight, type (quantitative, qualitative), number of intervals or values, list of intervals (only for quantitative descriptors), list of values (only for qualitative descriptors).[2]

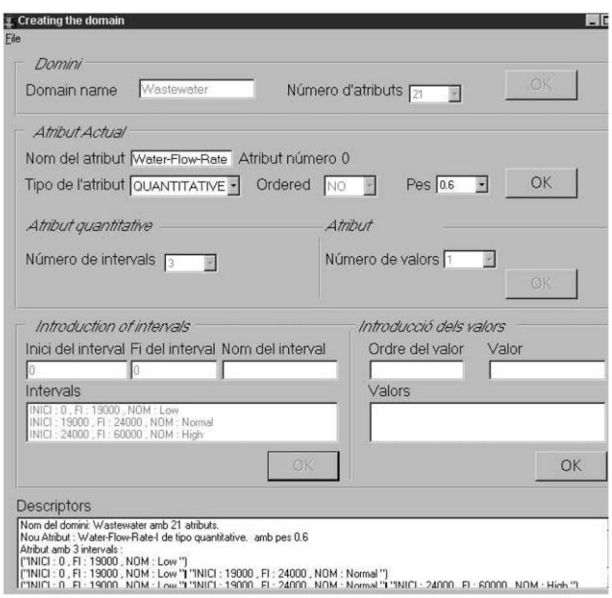

Fig. 5. Window for the creation of a domain in the CBRS of OntoWEDSS.

Actions are what the CBRS suggests as a reaction to a certain situation. Each action has the following structure: action identifier, action description, list of descriptors on which the action depends, action's formula, action's cut-off value. A *case library* represents the experience about the domain stored in OntoWEDSS. The case library stores only significant cases and not all of them. In OntoWEDSS, the CBRS represents the cases as a flat structure of non-ordered *descriptor–value pairs*. This is not a state-of-the-art representation schema and we recommend to change it to a hierarchical structure in future implementations. When the case most similar (e.g. 29.10.1999) to the actual problem (e.g. 14.12.1999) is retrieved, the CBRS *adapts the solution part of the case*. The solution is a list of actions (e.g. a change in the *waste flow of activated-sludge*, WAS) and each action has an associated formula. For instance:

$$\mathrm{WAS}_{\text{new-case}} = \mathrm{WAS}_{\text{retrieved-case}} \times \frac{\mathrm{Water-Flow-Rate-I}_{\text{new-case}}}{\mathrm{Water-Flow-Rate-I}_{\text{retrieved-case}}}$$

$$\mathrm{WAS}_{14.12.1999} = \mathrm{WAS}_{29.10.1999} \times \frac{\mathrm{Water-Flow-Rate-I}_{14.12.1999}}{\mathrm{Water-Flow-Rate-I}_{29.10.1999}} = 1100 \times \frac{24000}{22000} = 1200$$

This value (1200) will be the value of the suggested actuation, in the case of selection of CBR's diagnostics.

2.3. Ontology

The use of an *ontology*, *WaWO*[3] (Ceccaroni et al., 2000), helps to model the wastewater treatment process, standardizing the vocabulary and paying special attention to the management of the qualitative knowledge, that is, the environmental information on micro-organism presence. WaWO has been designed and built following current mainstream ideas about ontology construction, and is a hierarchically-structured set of terms and relations describing the domain of wastewater treatment (see Fig. 1). It is the manifestation of a shared understanding of the wastewater domain that is agreed among a number of experts in environmental and chemical engineering. The introduction of this agreed-upon ontology in the domain of wastewater treatment facilitates: (1) an accurate, effective *communication and sharing* of meanings, which leads to benefits such as knowl-

[2] A translation of the GUI from Catalan to English is under way.
[3] *WasteWater Ontology* (WaWO) availability: http://www.kslsvc.stanford.edu.

edge reuse, (2) an advancement in environmental technologies for the management of biological and biochemical processes, and (3) an enhancement of the knowledge on the specific microbial ecology of environmental processes developing in treatment plants. Even though WaWO was designed on the basis of the specification of a few particular plants, the knowledge which it embodies is valid for any treatment plant of the same class.

To build the *ontology*, after an exhaustive bibliographic research and taking into account the advice of other ontology engineers, we decided to adopt the Ontolingua formalism and development environment. Ontolingua is a complete work environment, which includes a specific language and an on-line editor (accessed through an HTML interface). The meta-ontology used by Ontolingua is the Frame Ontology which is used by default by all the ontologies being built on the server. The objects of the wastewater domain are represented as classes and individuals. In more general terms, these classes and individuals, as well as functions, relations and axioms, are all frames (in the context of the meta-ontology). An excerpt of the resulting Ontolingua-package-LISP source code is the following (WaWO-wastewater.lisp):

```
(Define-Ontology
WaWO-Wastewater
 (HPKB-Upper-Level-Kernel-Latest Frame-Ontology)
 “This is a specific ontology of …”

(In-Ontology (Quote WaWO-Wastewater))
;;; Compact
(Define-Individual Compact
  (Structured-Of-Floc))

...
;;; %-BOD-Removal
(Define-Class %-BOD-Removal (?X)
 :Def (Calculated-Descriptor ?X)
:Documentation “BOD percentage
  removed after primary, secondary
  or overall treatment.”)

…
;;; Foam-Presence-Aeration-Tank
(Define-Function
 Foam-Presence-Aeration-Tank (?Frame)
 :-> ?Value “Synonyms: ESC-B (cat)”
 :Def (And (Foam-Presence ?Frame)
  (String ?Value)))

…
```

With respect to the translation of WaWO from Ontolingua to LISP, we had to define or adapt a few basic routines to integrate the ontology into a standard *Common LISP object system* (CLOS). The most important procedure is a class-defining macro, which allows the generation of the hierarchical structure. A number of other functions and macros permit to keep track of defined instances.

2.4. Decision-support

With the decision-support task, we exploit the available data and information to provide: (1) active support about execution monitoring of RBES and CBRS, and (2) concrete action suggestions, such as “*Change Sludge-Recirculation-External to 120” or “Destruction of filaments via chlorine addition” or “Addition of inorganic coagulant”*. For example, when several factors may contribute to the need for a control action, the operator is warned of the problem and the diagnostics of the reasoners is presented to him. A course of action is proposed; the operator reviews the situation and determines the actuation. In other words, the decision-support task consists of gathering the integrated diagnostics of reasoners and ontology, activating the WaRP planner (Ceccaroni and Robertson, 2000) and selecting an actuation. Eventually, the output of the system is represented by statements about actions to be taken or statements to support a human manager’s decisions, in order to maintain the correct operation of the plant. The user is always given the possibility to ask the system for explanation about the results of RBES and CBRS.

3. Ontologies and reasoning

The WaWO ontology used in the system is specialized in the wastewater domain. In WaWO, for example, *Storm* is an *Operational-Problem*, *Bacterium* is a *Wastewater-Biological-Living-Object*, and the only *Metazoans* represented are *Nematode* and *Rotifer*. The use of the ontology for reasoning is completely experimental and there are no other documented studies in a similar context. The hierarchical organization of categories of WaWO is expressed in the Ontolingua knowledge-representation language, and KIF axioms are used for answering queries, language analysis and general reasoning. In Fig. 6, an example of the reasoning with the ontology is depicted. It can be partially read, using the terminology of Sowa (2000), as a sequence of occurrences. Simple rectangles are role categories (or phenomenon categories, or classes) and are always part of concept hierarchies; circles are relations. In the example, *Filamentous-Bacteria* is what causes (i.e. is the effector of) the *Filamentous-Bacteria-Excessive-Proliferation* occurrent. Being the effector part of taxonomic (*Microfauna* branch) and operational (*Microthrix-parv-*

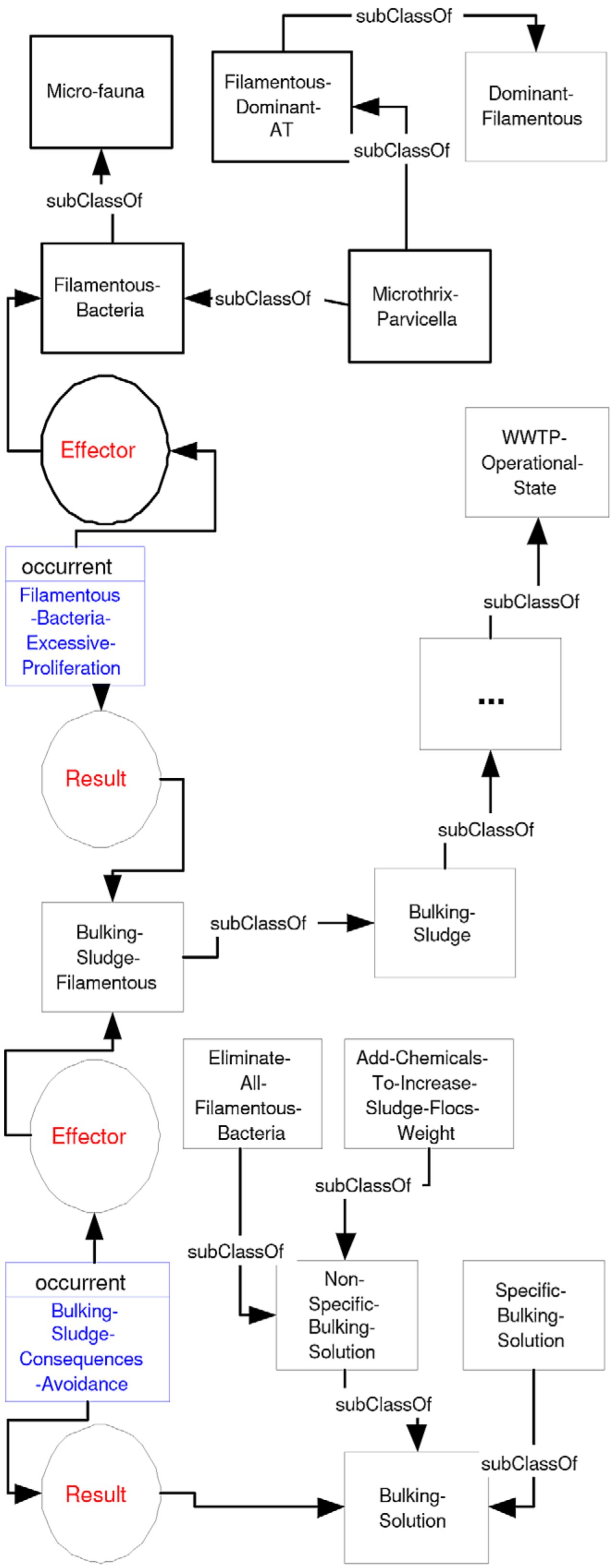

Fig. 6. Reasoning with the ontology.

icella branch) semantic structures, the occurrent can then be linked both to a class of micro-organisms and to the *Filamentous-Dominant-AT* state of the WWTP. The result of the occurrent is the *Bulking-Sludge-Filamentous* situation, which is itself the effector of the *Bulking-Sludge-Consequences-Avoidance* occurrent. The result of this last occurrent is the search for a *Bulking-Solution* within its concept hierarchy. The axioms (which are essentially rules) guide the search towards a specific or a non-specific solution, according to the original effector. The final result can be, for instance, that the presence of a certain amount of filamentous-bacteria triggers the addition of certain chemicals to the wastewater. Even if it is true that the same result can be achieved through a complex RBES, the very structure of the ontology guarantees much greater readability, easier consistency-checking, and reuse of knowledge in a way that is becoming a more and more recognized standard. Furthermore, the ontological knowledge representation is very web-oriented and could be useful in a future scenario of knowledge sharing over the Internet.

4. How the system works?

In this section, the use of OntoWEDSS is described in detail. OntoWEDSS has not been deployed in a full-scale facility; its application and evaluation, however, are based on data from full-scale facilities. The ontology, in its current state, is fully reusable if transferred to any real facility. The same is true for the CBRS, while in the case of the RBES, only the rules of the upper layer (see *RBES* in Section 2.2) may be reused without modification. OntoWEDSS's interface allows the user to communicate with the RBES and CBRS in a friendly way, via the selection of options from menus and buttons. The interface has three main functionalities: (1) the introduction of the data of the problem in question, (2) consulting the RBES about the state of facts or rules, and (3) asking the user to confirm actions or about data values. The use of OntoWEDSS can be conceptually subdivided into two parts: (1) *domain modeling*, which facilitates search-space handling, result presentation and domain organization, and (2) *execution*. We will now describe how OntoWEDSS works through examples, which have been slightly simplified, reflecting a real situation.

4.1. Domain modeling

4.1.1. Domain definition and modification in the CBS

In the case-based reasoning system, we can create a completely new domain. This is done by defining all descriptors for the domain, e.g. Name: *Water-Flow-Rate-I*; Type: quantitative; Weight: 0.6; Number of intervals: 3; Intervals: *Low* [0, 19 000], *Normal* [19 000, 24 000], *High* [24 000, 60 000]. And so on for all descriptors of the domain. As an alternative to creation from scratch, a domain can be loaded, visualized and then, if necessary, edited.

4.1.2. Domain definition and modification in the RBES

In the rule-based expert system, we can also create a completely new domain or reuse an existing one. Additionally, rules have to be designed or reused from an existing KB. Designing a new rule is done defining a few parameters, e.g. Name: *bsf-1* (first rule for bulking-sludge-filamentous); Module: *bulking*; Invocation conditions: (*pre-alarm-filamentous-bulking* is true) *and* (*SVI - AT* > 140) *and* (*filam-dominant-at* is not *none*); Conclusion: *Bulking-Sludge-Filamentous* is true. And so on for all rules.

4.1.3. Domain definition in WaWO

WaWO is the foundation of the modeling of the domain. It serves search-space handling in the following way:

- Expanding the search: querying with similar concepts (using the boolean *or*);
- Reducing the search: querying with more specific concepts;
- Searching cross-lingually: expanding the search using available translations of the terms (in the case where non-standard categories are used).

It is possible to load the ontology, but it is not possible to edit the ontology directly from the user interface. It is instead possible to do it using the routines described in Ceccaroni (2001).

4.2. Execution

4.2.1. Execution of the CBRS

The execution consists of loading the domain, loading the case library, introducing a new problem, e.g.: ("Day" 5, "Month" 11, "Year" 2000, "Hour" "01-00 pm", "Water-Flow-Rate-I" 22166, "TSS-1" 276.00, "COD-I" 739.00, "TKN" "?", "TSS-P" 88.00, "COD-P" 556.00, "MLSS-AT" 2660.0, "SVI-AT" 139.1, "SRT-AT" 0.00, "F-M-AT" 0.00, "Filam-Dominant-AT" "?", "DO-AT-1" 2.45, "DO-AT-2" 2.44, "TSS-E" 27.00, "COD-E" 96.00, "TN-E" "?", "Diagnosis" -, "WAS" -, "RAS" -, "Air-Flow-1" -, "Air-Flow-2" -). The CBRS tries to retrieve a similar case and, if no case is retrieved with a similarity greater then *b*, the 5 "most similar" cases are shown to the user anyway.

4.2.2. Execution of the RBES

The execution consists of loading the domain, loading the rules, introducing a new problem. Conclusions appear in the *Deduced conclusions* window, e.g. (*Bulking-sludge-filamentous* true 1.0). The trace of the reasoning appears in the *Trace* window. The trace is always a path in one or more decision trees (see Fig. 3).

4.2.3. Execution of WaWO

After the execution of the CBRS and the RBES, OntoWEDSS automatically loads and compares the results. Then, the WaWO ontology is instantiated and used to prove statements (see Section 5).

5. Performance

In the industrial domain, when a new software system is proposed, it is important to prove that it is better than the previous ones, through a statistical evaluation. The contribution of our research, nevertheless, is more about the introduction of a new paradigm, rather than a new software. Additionally, in environmental–industrial mixed domains, such as wastewater treatment, evaluation is problematic and complex for a number of reasons:

- *Lack of benchmarks.* There is a lack of benchmarks that work, with regard to wastewater management. Now that so many WWTPs are in operation, an accurate benchmark is critically needed. This situation does not allow an accurate quantitative measure of management improvement.
- *Different descriptors*. For our research, 170 descriptors are available. Experts helped us to select 21 of them to use in CBR and RBR. The chosen descriptors are the *most* relevant in experts' practice and experience, but not all systems use the same descriptors.
- *High percentage of missing values*. In real-world environmental applications, we are usually faced with a high number of instances with missing descriptor's values. These instances are often not suitable to be correctly labeled.
- *Multiple labels.* In wastewater treatment domain, it is possible to assign more than one *Diagnosis* label to a state of the plant (e.g. *Bulking-Sludge*, *Underloading* and *Rising-Sludge*), ordered according to importance. This situation makes the evaluation of diagnostics more difficult. We chose to work with just one label per instance to ease the validation process and this degrades, in part, CBR's performance.

Due to these limitations, the OntoWEDSS system was not evaluated against other systems, but only against previous versions of the system which do not include the ontology. However, as explained earlier, performance is not a primary concern of this research, which focuses on the introduction and integration of knowledge-representation techniques. The *evaluation* of the system was limited to the most representative problematic situation that is possible to come upon in wastewater treatment (the presence of bulking sludge due to filamentous micro-organisms) and was carried out just as a demonstration of the great potential of introducing ontologies. The objective of the evaluation was to assess the performance of the various paradigms and of the whole system when they react to a specific problem. In the three experiments carried out, the *successful diagnosis* coming from the system without the ontology ranges from 60% to 73% and impasse situations correspond to a set of ten instances (out of 57). When the system operates with the WaWO ontology, these results improve. This improvement is necessary, because we augmented the system with an additional module, so that the minimum performance equals the maximum performance of the previous one. The improvement in diagnosis is due to the following two circumstances: (1) WaWO activates when an impasse situation has been reached and takes this into account, sometimes forcing a weaker solution; (2) WaWO usually has at its disposal additional information about micro-organisms that was not used in the earlier evaluation because the RBES and the CBRS are not designed to deal with it. The final successful diagnosis coming from the system with the ontology ranges from 73% to 100%. We acknowledge that, in one of the experiments, the ontology could not improve the performance, which was 73%.

In the following part of the section an account of WaWO role in the analysis of the all ten impasse situations is given. The two basic descriptors which are used in bulking diagnosis are *SVI-AT* and *Filam-Dominant-AT*. In general, WaWO not only tries to detect filamentous-bacteria excessive proliferation, but also offers a specific actuation strategy, according to the identified bacteria. In case the bacteria causing *Bulking-Sludge-Filamentous* are not determined, a non-specific solution (e.g. adding chemicals to increase the weight of the sludge flocs or eliminating all filamentous-bacteria) is offered to avoid the consequences of bulking sludge.

- (“Day” 25 “Month” 10 “Year” 1999)
(“Day” 23 “Month” 5 “Year” 2000)
(“Day” 25 “Month” 8 “Year” 2000)
Microthrix, the value of *Filam-Dominant-AT*, is a subclass of *Filamentous-Bacteria*. A relation connects *Filamentous-Bacteria* to *Bulking-Sludge-Filamentous*, which is in WaWO a subclass of *WWTP-Operational-State*, that is the category used for diagnosis expression. Nineteen other relations connect, according to the particular dominant micro-organism (*Microthrix* in this case), the *Bulking-Sludge-Filamentous* class to a specific *Actuation*.
- (“Day” 16 “Month” 5 “Year” 2000)
(“Day” 18 “Month” 8 “Year” 1999)
SVI-AT is a subclass of *Sludge-Volumetric-Index*. A relation connects *Sludge-Volumetric-Index* to *Bulking-Sludge*. Another relation connects *Bulking-Sludge* to a non-specific *Actuation*: the destruction of filaments via chlorine addition up to 20 mg/l. If this actuation is not available for any reason, another relation connects to a second non-specific actuation: the increase of the characteristic weight of flocs via inorganic coagulants addition, such as lime or ferric salts.
- (“Day” 12 “Month” 9 “Year” 1999)
Same as 25.10.1999, with *Gordona* instead of *Microthrix*.
- (“Day” 30 “Month” 6 “Year” 2000)
(“Day” 26 “Month” 7 “Year” 1999)
(“Day” 11 “Month” 8 “Year” 2000)
(“Day” 7 “Month” 9 “Year” 2000)
No diagnosis recommendation.

5.1. Why OntoWEDSS is better than earlier systems?

As seen in Section 1.2, the earlier systems of reference for OntoWEDSS are DAI-DEPUR (Sànchez-Marrè, 1995) and BIOMASS (Comas, 2000). OntoWEDSS improves these systems in several ways, but mainly by the introduction of an ontology and the addition of a diagnosis integrator. Indeed, OntoWEDSS presents a novel integration between KBSs in a real-world application. The integration happens mainly at the diagnosis level, where the results of RBR and CBR systems are compared before providing information for the decision-support task. A system of priority is established among the KBSs as well as the cases in which WaWO is called. Unlike DAI-DEPUR and BIOMASS, in OntoWEDSS, a semantic integration of information exists. In fact, a problem that goes beyond syntactic integration (dealt with by BIOMASS and in part by DAI-DEPUR) is the mapping of semantics of terms from different information sources (such as different WWTPs), even when these terms have been expressed using the same syntactic structures. For instance, even when two applications use the same language as their interchange format, how can we be sure that the same words in their vocabularies mean the same things? The WaWO ontology is an instrument to solve semantic problems of this kind. Being WaWO stored in the well known Ontolingua Server, its sharing is easy and the knowledge-representation formalism is standard. Moreover, WaWO can be translated into several implementation languages thanks to Ontolingua translators. The lexicon and semantics of WaWO are as standard as possible, synonyms are shown in the documentation and there are no hidden assumptions. Solving part of the existing terminological confusion, OntoWEDSS matches more properly the domain needs. Furthermore, the hierarchical structure and the axioms of WaWO help to diagnose the situation in the case of an impasse of the other KBSs, allowing reasoning on different levels of abstraction.

While in DAI-DEPUR there is no modeling of wastewater microbiology, in the OntoWEDSS system, the microbiological component is modeled by the ontology, and this opens new possibilities of search and inference in the process of WWTP control. An identification of the most common micro-organisms and a comparative study of micro-organism communities of different treatment-plants have been carried out to understand the influence of biological variability at a geographical level. Then, a set of microbiological descriptors have been selected to be used by WaWO, together with the standard physical and chemical ones. Also, through the definition of ontological relations, we represented two kinds of real-world cause-effect relations: (1) association of micro-organisms to the problematic situations that they cause and (2) association of the actual state of the plant to the actions that need to be performed in order to reach the normal state from that actual state.

6. Conclusions

We presented a new architecture for environmental decision-support systems and developed a prototype (called OntoWEDSS) for the domain of wastewater treatment. The main new characteristic of OntoWEDSS with respect to previous similar systems is the integration of case-based and rule-based reasoning with an ontology, WaWO, for the representation of the domain and for reasoning. This integration improved the modeling of the information about wastewater treatment processes and resolved existing *impasses* in the reasoning cycle. We presented an ontological representation of two kinds of cause-effect relations: *micro-organisms↔problematic situations* and *state of the plant↔suggested actions*. We also used the ontology to improve the communication among different elements and agents of an environmental decision-support system, thus reducing ambiguities. Thanks to the use of this ontology and exploiting all the data on activated-sludge, the OntoWEDSS system aims to go a step further in completing the comprehension of micro-organisms living in treatment plants and in using this knowledge for a better management.

Acknowledgements

This research is supported by ATEAM (IST-1999-10176) and Agentcities (IST-2000-28384) European projects, and the CICyT TIC-2000-1011 Spanish project. We wish to thank Jonathan Dale, Dominic Greenwood, Richard Rabbat (Fujitsu), Steve Willmott (KEMLg-UPC) and the anonymous reviewers for the feedback after having read a preliminary version of this paper.

References

Ceccaroni, L. Integration of a rule-based expert system, a case-based reasoner and an ontological knowledge-base in the wastewater domain. In Proceedings of ECAI 2000 - W07: Binding Environmental Sciences and Artificial Intelligence (BESAI2000), pages 8.1-8.10, Berlin, Germany, 2000.

Ceccaroni, L., 2001. OntoWEDSS—An Ontology-based Environmental Decision-Support System for the management of Wastewater treatment plants. Ph.D. Thesis, Universitat Politècnica de Catalunya, http://www.lsi.upc.es/~luigic/thesis.pdf.

Ceccaroni, L., U. Cortés, and M. Sànchez-Marrè. WaWO - An ontology embedded into an environmental decision-support system for wastewater treatment plant management. In Proceedings of ECAI2000 - Wo9: Applications of ontologies and problem-solving methods, pages 2.1-2.9, Berlin, Germany, 2000.

Ceccaroni, L., Robertson, D., 2000. WaRP - A Reactive Planner integrated in an environmental decision-support system for Wastewater treatment plant management. In Proceedings of ECAI 2000, Berlin, Germany, 491–495.

Comas, J., 2000. Development, implementation and evaluation of an activated sludge supervisory system for Granollers WWTP. Ph.D. Thesis, Departament d'enginyeria química agrària i tecnologies agroalimentàries, Universitat de Girona, Girona, Spain.

Cortés, U., Sànchez-Marrè, M., Sangüesa, R., Comas, J., R-Roda, I., Poch, M., Riaño, D., 2001. Knowledge management in environmental decision support systems. AI Communications 14, 3–12.

Fensel, D., Horrocks, I., Van Harmelen, F., Decker, S., Erdmann, M., Klein, M., 2000. In Dieng, R., (Ed.), OIL in a Nurshell, pp. 1-16.

OntoWeb project, 2002. OntoWeb—Technical Roadmap v2.0. Deliverable 1.1.2 of IST Project IST-2000-29243 OntoWeb, November 2002.

Sànchez-Marrè, M., 1995. DAI-DEPUR: an integrated supervisory multi-level architecture for wastewater treatment plants. Ph.D. Thesis, Software Department, UPC, Barcelona, Spain.

Sànchez-Marrè, M., Cortés, U., Lafuente, J., R-Roda, I., Poch, M., 1996. DAI-DEPUR: an integrated and distributed architecture for wastewater treatment plants supervision. Artificial Intelligence in Engineering 10, 275–285.

Sowa, J., 2000. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole.

Stuckenschmidt, H., 2000. Using OIL for Intelligent Information Integration. In Proceedings of ECAI2000 - W09: Applications of ontologies and problem-solving methods, pp. 9.1-9.10, Berlin, Germany.

Swartz, A., Hendler, J., 2001. The semantic web: a network of content for the digital city. In: Proceedings of the Second Annual Digital Cities Workshop, Kyoto, Japan,

van Heijst, G., Schreiber, A.T., Wielinga, B.J., 1997. Using explicit ontologies in kbs development. International Journal of Human-Computer Studies 46 (2-3), 183–292.

W3C, 1998. PICS Signed Labels (DSig) 1.0 Specification. W3C Recommendation, May 1998.