Expert System for Monitoring Pollution of Rivers and Lakes

Claudio G. Bernardo, Diego K. Souza, Jackson G. F. Nascimento, Ricardo P. C. Kaipper, and Thaigo C. M. Alvim

Abstract—This research presents the state of the art of expert systems applied to one solution for monitoring and control of pollution of Brazilian rivers and lakes, because of the many cases of indiscriminate pollution of its rivers and lakes and it comes to affect people's lives and biodiversity as a whole. This application generates reports and supports traceability creation and cross-reference information to search for possible causes and appropriate solutions to the problem raised. Its main contribution is supporting the government and its human specialists how to deal with this great environmental problem using computers as specialists and help them choosing what decisions to make.

Index Terms—Expert system, control of pollution, rivers and lakes monitoring, decision make support.

I. INTRODUCTION

One of the biggest environmental problems today is the indiscriminate pollution of rivers and lakes throughout the world. According [1] these freshwater sources bring a huge benefit to society and have great importance to the organization of geography, with social and economic value. In areas of high demand, especially in metropolitan areas, these river structures are suffering negative consequences due to urbanization. The daily dumping of waste into its waters make those rivers and lakes eventually become open sewers, being in many cases treated as byproducts of urban society. According to [2] the program of the United Nations Environment (UNEP) of 1990 related the main groups of marine pollutants that impact coastal areas and oceans worldwide, the sewage, persistent organic compounds, radioactive elements, heavy metals, nutrients containing nitrogen and phosphorus, hydrocarbons, suspended solids (sediment movement) and solid waste. In an overall assessment, the two groups of pollutants that showed most impact on the ecosystems of the Brazilian coast were domestic sewage and persistent organic compounds. The sewer, due to the high volume and because they are often dumped almost untreated, contribute to the contamination of the domestic basins. This contamination is difficult to supervision and control due to the large land area of the country and the immense formation of veins, rivers, springs and tributaries.

In view of the great difficulty in controlling what is dumped into rivers and lakes, the proposal to develop an intelligent system to perform verification and control of the river system and to identify the main causes and the areas most affected by these attacks is showing to be a useful tool. The installation of a system that analyzes the status of the water in order to maintain updated information about the watershed status, and that checks the possible cause of contamination is of great importance for the country. The analysis will be essential in the system, because depending on the situation raised and the data collected in rivers and lakes, the agent could cross information and look forth the causes of each source of pollution, providing then a solution, contributing to the process of decision making when developing a plan of decontamination and preservation.

II. EXPERTS INTELLIGENT SYSTEMS

An Expert System (ES) is understood as a computerized system that uses knowledge of a specific area of application and complex to act as an expert consultant. It brings together all the elements of the decision process that a human expert can provide. Consists of a knowledge base and software programs that perform inferences on knowledge and transmit answers to the questions of a user. According to [3] the ES is defined as a computational system aimed at bringing together all elements of the decision process of a human expert. Gather expertise on specific areas and fields of knowledge is a prerequisite and this knowledge should be broad and well organized for easy reference made by a user who is not an expert. Even if human experts and expert systems perform the same task, they both vary critically. In this context, an ES as well as an expert, can or not reach a solution, or even get a distorted solution. According to [4] the system can go wrong, but its error or no response stems from certain circumstances justified by the system itself.

For the construction of an ES, researchers of [5] claim that the computer needs to process human knowledge and for that reason this knowledge must first be represented or modeled. The knowledge of human experts is stored in a knowledge base which can be represented by means of rules of production systems, semantic networks and frames, among other forms. According to [6] the diversity of fields such as medicine, engineering, education and business support, using solutions based on expert systems, enabled the exploration of various categories of problems. Among these categories are: Interpretation, Diagnostics, Monitoring, Prediction, Planning, Design, Debug, Repair, Instruction and Control.
A. Components and Architecture of an Expert System

The basic components that interrelate in an expert system include a System of Knowledge Acquisition, Knowledge Base, Blackboard and Inference Machine [7]. Reference [8] states that the acquisition of knowledge in humans is taken through the learning and by the point of view of a system, this process may be defined as changes in the system that allow to redo the same tasks more efficiently in the future. This definition characterizes the ability to acquire knowledge and develop a learning process by an Expert System. Reference [8] states that learning becomes possible relying on experience and taking into account the compression of the data sets. A component of an ES is the key to build a knowledge base and allows the learning process by the system, so that the interaction performed by the system in the process of knowledge acquisition, compressing the experiences recorded in the base data, is what ensures that learning and the success of the support system. A knowledge base contains facts about an area that has a very specific topic (technical knowledge about the specified science) and heuristics (principles votes) expressing their procedures of reasoning of an agent in relation to the specified subject. Reference [9] states that the knowledge base consists of the rule base and working memory, as well as blackboard, inference engine with forward chaining, backward chaining and production rules. According [10] working memory is a mean of representation and control of the knowledge base using rules group sin dependent of name. In the prototype presented in this research the production rules may take other values or even more of a value to represent a result, in the case of multivalve variables. These elements are inserted and are interconnected within a generic architecture which can be seen in Fig. 1.

A significant knowledge base regarding the subject or problem area, besides having at disposal the provision of this data set and information in the form of organized data structures, is crucial for an ES to solve a problem efficiently. According to [7] the representation of knowledge is suggested as the formalization of the system’s knowledge, and to make this possible some techniques that allow modeling knowledge more efficiently are used, the main ones being the logic of propositions and predicates, rules production, semantic networks and frames.

B. Applications of Expert Systems

The basic components of an E S that are presented in Fig. 2 can be applied to numerous tasks, usually focused on exploitation of problems and proposed solutions or diagnoses. Its activities include management decisions, troubleshooting, operation, maintenance/programming, design/configuration, selection/classification, and monitoring/control process.

Practices where you can apply an ES are the most diverse and by a demand for verification, identification, diagnosis of causes and appropriate solutions they meet the needs of the context of implementation of an intelligent agent for verification of rivers and lakes. From all the information gathered and observing the needs and complexity of implementation area for the proposed objective was chosen using the shell Expert SINTA [11] for preparation of an ES as rule-based tool for the proposed development.

Rules can be understood as the set of rules composing clauses and that run the system knowledge. According [7] each rule consists of a mandatory provision in its antecedent clause part and mandatory in the consequent part. Any number of clauses linked by logical connectives ("\&\&" or \"\>|\") may appear in the antecedent part of a clause.

C. Shell Expert SINTA

The Expert SINTA, according [5] is a tool for generating expert systems through the use of artificial intelligence techniques within the computing environment. The tool aims at simplifying the task of implementing expert systems, through the use of a model of knowledge representation based on production rules and probabilities and also the use of a shared inference machine, the use of screens and menus automatically constructed, the probabilistic treatment of production rules and the use of context-sensitive explanations of the knowledge base that are modeled and types of ideal models in classification problems. Some examples of ES with these characteristics include diagnoses of carbon-14 dating systems, systems used in archeology and configuration of computer networks, among other systems. The Expert SINTA shown in Fig. 3 was created for knowledge analysts who wish to implement a knowledge base for your domain, so that this does not require the analyst programming knowledge. To make the management of a knowledge base, its creator must provide information about your system variables, rules, questions and goals.

III. INDISCRIMINATE POLLUTION OF RIVERS AND LAKES IN BRAZIL

The biggest challenge of this millennium is to avoid water shortages. According [12] studies in this century indicate that approximately two billion people face water shortage in the world. Continents most affected by the shortage are Africa, Central Asia and Middle East. Between 1990 and 1995, the need for water doubled compared to the world population, due to their high consumption in industrial activities and agricultural areas. It is now known that only 2.5% of the planet’s water is fresh and less than 0.1% are in areas accessible to humans.

The degradation of rivers and lakes is mainly caused by pollution and contamination by pollutants and sewage discharged without proper treatment. One of the main culprits in these cases is synthetic organic compounds, which for decades have been inserted into the environment. These compounds are virtually non-degradable in the environment, with a final repository in the deep ocean. The impact of these
substances is so high that, according [2] the United Nations program for the prevention of marine pollution from land-based activities provided the ban of the use of all chlorinated organic compounds used as solvents. On the other hand, the problem with the drains is showing to be much more complex, once the wastewater treatment requires great investment and drain discharge is something that cannot be prevented or replaced as occurs with solvents. A deficiency in these sewage treatment reflects an indifference to sanitation, as well as the quality of water provided to the population.

Brazil, as a country with one of the largest watersheds in the world, suffers from the problem and the lack of control and monitoring of its water, which is difficult to make an analysis of the cause / effect of contamination of these basins is a major aggravating factors. In view of this problem the Brazilian government finds great difficulty to monitor and ensure the correct use of their basins. Another factor that hinders the enforcement of the waters is the rapid urbanization of the country without planning, control and structure.

Fig. 2. Basic components of an expert system.

Fig. 4 shows a table with the degree of seriousness of Brazilian environmental problems. Today’s major assaults on the ecosystem come from the population, which has no awareness of the conditions and needs of the ecosystem and this causes large amounts of urban waste such as plastic bottles, organic wastes, detergents, furniture among others, dumped in rivers and lakes daily by this population. Another factor that contributes to this aggression is the dumping of sewage without any treatment in these waters. Some countries deal with very similar problems and seek appropriate solutions to get the rehabilitation of rivers and lakes to restore the balance of the ecosystem in that region. Measures adopted for the remediation of its waters first started by identifying the main causes of this imbalance and then to analyze what have to be done to correct and clean effectively these waters.

Due to its large area and abundant territorial extension, an execution analyzes and identification of related problems of the pollution of Brazilian rivers and lakes are very difficult. It is necessary that researchers increasingly require the use of new technologies as computer systems and techniques for more effective combat and identification of zones of water pollution. The use of intelligent systems in order to support the process of decision making of plans that aim to contribute to the environment is already something that has been used throughout last decade. Artificial intelligence had great participation on new technologies applied to aid decision making on fighting pollution and controlling the waste of waters, these systems can provide higher mobility and ease unattended control of environmental assaults.

The Standard 307 published by [13] regulates in Brazil the correct procedure to perform with respect to medical waste and the destination to be given to them. Divide the hospital waste in waste groups based on their characteristics and each group is set to the correct form for storage and disposal as well as packing and packaging for these residues. The classification of these elements is divided in special waste, common waste and infectious waste. According [14] are constituted of septic wastes that contain or may contain potentially pathogenic organisms and are produced in health services, such as hospitals, clinics, laboratories, pharmacies, veterinary clinics and health posts. Despite the great importance of this matter for the maintenance of water quality and environmental balance, analysis of hospital waste dumped into freshwater sources was not considered in this research.

IV. EXPERT SYSTEM MONITORING BRAZILIAN RIVERS AND LAKES

The methodology used in this research is directly linked to the ability to check and identify common problems that occur in Brazilian basins. The project proposes to use a system based on expert system to verify the conditions and changes in rivers and lakes in several parts of his compositions application. It is based on the comparison of data quality analysis of points of rivers and lakes sent for the intelligent agent by direct insertion into the system, on the fill of forms that collect information, associating and generating a framework of possibilities for the situation which will be analyzed to generate a possible solution to the problem as well as pointing out the probable cause of the disorder. If it becomes clear that a solution should be used to assist in the decision making process. The rules and features raised were
formulated by websites of companies specialized in water treatment and modifications, and adjustments are acceptable to the extent that new forms of verification and ecological control is usable.

Reports and Interfaces Used in Research Elaboration

The work of the Expert rules SINTA occurs by completing a form where each question assigns a value to a variable. The form was divided into nine levels of analysis, enabling the verification of results at each level, allowing cross results between levels getting higher precision in the final result. The analysis is performed through multiple forms with fill options that indicate the material properties as sample. These accept multiple responses for each test and each has a composition of statements which must be marked according to the analysis of the sample and the environment. The generated reports were:i) Report of environmental analysis - checks conditions and provides additional information on the situation in the region where the sample was collected. ii) Report of microbiological analysis - checks for the presence of *scherichia coli* in the sample; iii) Report of cyanotoxinsanalysis - checks for the presence of cyanotoxins in the sample; iv) Report for potability pesticides analysis - checks if the amount of pesticides is within the permissible maximum values; v) Analysis Report for potability inorganic substances - verifies that the amount of inorganic substances is within the maximum values; vi) Report for potability organic substances - verifies that the amount of organic substances are within the permissible maximum values; vii) Analysis Report for potability inorganic substances - verifies that the amount of pesticides is within the maximum values; vi) Report for potability inorganic substances - verifies that the amount of organic substances are within the permissible maximum values; vii) Report for potability disinfectants analysis - checks if the amount of disinfectant is within the permissible maximum values; viii) Report of the organoleptic analysis - checks if the amount of disinfectant is within the permissible maximum values; ix) Report of radioactivity analysis - checks if the amount of radioactivity is within the maximum permissible values.

Variables Used in Prototype Elaboration

In Expert SINTA all variables were defined as one valued, multivalued and numerical. The one valued variables are those that accept only a value to be assigned to each of them, may be a boolean value with true or false or even a name or numeric value set. The amount of 29 variables in the proposed application to effect proper control of water quality were used. The multivalued variables are those that accept more than one value, since the same type (boolean, integer, decimal, etc.) and serve to allocate and define characteristics of an element of question. These variables and the analysis reports (with the exception of the analysis of the environment report) make up the body of environment analysis, microbiological and physical / chemical water. Their values were obtained from tables and indicators published in the order of Brazilian Health Ministry [15] and from adapted values of procedures published by [16] is considered as the ideal values for human consumption and environmental balance. The application rules are created considering the provisions of this ordinance and have been applied in the questionnaires whose answers were filled manually.

The check determines whether the water is fit for consumption or not, proper to dwell life or not, whether are contaminated with any substance harmful to animal life or not and if water and soil are contaminated or not. The check even responds with an odds ratio with the potential responsible for contamination of a stretch as well as possible solutions to each problem encountered within the possibilities. In the proposed prototype 163 rules for prototype building and completion of the analysis of water were used.

Trust Factor of an Expert System - CNF

The trust factor of an information (CNF is the acronym in Portuguese) in the shell Expert SINTA uses an approach based on possibilities, assigning in a generalized form factors of certainty about the rules. This kind of approach is based on the human trait of not being deterministic, since there is no absolute certainty about the information provided and therefore a percentage of confidence is assigned to this information. In an example adapted from [17] you can check the construction of a factor of certainty in the shell:

- If it will rain = yes
- Then do not go to the beach = yes, with 90% CNF.

In this case if the premise is raining today = yes means 100% since it was not informed, then do not go to the beach = yes is 1.0 × 0.9 = 0.9 = 90%.

If the premise raining today = yes is 80%, the confidence level of completion "I'm not going to the beach = yes" will be 0.8 × 0.9 = 0.72 = 72%.

Where the assumptions submit connectors "and" the degree of confidence is given by multiplying the degrees of confidence of their respective assumptions of this rule.

Where the assumptions submit connectors "or" the degree of confidence will be calculated as follows: (CNF1 × CNF2) - (CNF1 × CNF2).

The application admits a minimum 50% confidence value to assume that inequity is true. In an ES, to define and find answers about a particular problem (represented by variables), defining objectives are necessary, which are defined as target variable. In the preparation of this research, only one target variable for each level of analysis was used: the multivalued variable "Properties Report" to level of physical / chemical analysis pointing out the elements that are irregular in the sample; the multivalued variable "Analysis Result" to represent possible signs of pollution and multivalued variable "EnvironmentalReport" to represent what measures should be taken to control pollution in the region.

Development of SEAQUA Prototype for Water Analyze

The chosen name for the knowledge base of the Expert System to analysis and display of patterns of contamination
of Brazilian rivers and lakes was SEAQUA (SE means IF in Portuguese and AQUA means WATER in Latin). The creation and management variables are fundamental to the construction of an expert system. Before you define and create the rules under development is necessary that all variables used and their values are created and released on Expert SINTA. Since the goal of a human expert consultation is to find the answer to a problem, the same holds for an expert system by computer, with the difference that it; the "problems" are represented by variables.

To consult the System SEAQUA, the consultant developed through menus of multiple (or only) choice, presented in Fig. 5. The degree of confidence is not required due to the selected values are related to empirical results from tests already carried out. This query is performed by just selecting as many boxes as necessary (checkbox on "Option"), and then selecting "OK". Not checking any box is also a valid option.

The user selects the desired options (more than one option can be chosen) by checking the appropriate checkbox, and then optionally informing the confidence level (when you do not have absolute certainty about a fact a percentage of confidence on the choice should be inserted) if not entered, the system automatically understand that the confidence level is 100%. In the case of SEAQUA system, the degree of trust is not required due to the selected values are related to empirical results from tests already carried out. A results window, which is shown in Fig. 6, is displayed upon completion of searches each objective and these windows tell the result found in the ES and the path taken to get to this answer.

In this report are listed the causes of pollution and measures which should be checked to contribute to the water quality of the region, where it is possible to conclude the fact that there are debris dumped in the waters of this region that cause pollution of these, as well as action to be taken, which is to educate the public through social / educational measures. The degrees of confidence presented at the end of the analysis are the calculation of all CNF generated by the results of the analysis of the properties, which represent the degree of certainty of action to be taken and consistency of facts addressed.

E. Main Rules Used by Prototype

A total of 163 rules were used in the design of the prototype SEAQUA. Twelve rules were considered major selected, but for the sake of space are presented in this paper only two:

Rule 1 - Coliform Contamination
IF:
Escherichia Coli= Present in the sample of 100 ml of water
THEN:
Microbiological Result = Indication of contamination by coliforms in 100ml of water result. CNF 100%
Microbiological Result = Evidence of imbalance found: CNF 100%
Micro-Indicator = Not CNF 100%
Properties Report = Microbiological. CNF 100%
Properties Report = Analysis Properties. CNF 100%
Properties Report = Escherichia coli in 100ml of water. CNF 100%

Rule 68 - Organic Substances
IF:
Ind-Pot (O) = No
THEN:
Potability Results (organic) = Level of Organic Substances above permitted values. CNF 100%
Potability Results (organic) = Evidence of imbalance found: CNF 100%
Properties Report = Organic Potability: CNF 100%
Properties Report = Analysis Properties: NFC 100%.

V. CONCLUSION AND FUTURES APPLICATIONS

Expert systems based on AI are useful tools in environmental control. With this vision, the SEAQUA prototype for analysis of water contamination was developed, to support the human expert in decision-making, analyzing the water conditions in the area and finding possible causes of the problem.

In view of the complexity and reliability that the information provided should offer were used as a basis of comparison indicators proposed by the Brazilian Health Ministry as the ideal values for human consumption and environmental balance. By means of realized tests and compared with indicators of pollution published by Brazilian government, was found that the SEAQUA system meets the intended goals, assisting in decision support experts in environmental analysis. The authors evaluate that using expert systems in environmental control is a good investment to ensure improvements in the control of the Brazilian sources waters.

The use of shell Expert SINTA led to the creation of the prototype SEAQUA. This use was adequate for its purpose, but other ways to design the ES can ensure the even bigger and better decision support systems of the institutions features, overcoming limitations of the shell. Next ES faced to environment will ensure more automation in the process of analyzing and constructing results.
Automating the collection of data for this solution can be improved in upcoming updates, given recent innovations in the way we collect data for analysis. It was concluded that with the evolution of technology, it is increasingly necessary to invest in intelligent applications that can provide better performance in controlling the ecosystem, increasing assertiveness in applications, to ensure the provision of information coming to reality, for analyzing, classify and treat problems affecting the biosphere.

REFERENCES