Ontological Knowledge Base Development for the Business Process Control in a Wastewater Treatment Plant

Chernyakhovskaya, L.R
Computer science and robotics department,
Ufa State Aviation Technical University
Ufa, Russia
lr_chern@imail.ru

Shkundina, R.A.
Computer science and robotics department
Ufa State Aviation Technical University
Ufa, Russia
r.shkundina@mail.ru

Muelle, J. A.
Institute of Programming Structures
and Data Organisation
University of Karlsruhe
Karlsruhe, Germany
muelle@ira.uka.de

Ugryumov, M.S.
Bank's Soft Systems Ltd,
Moscow, Russia
m.ugryumov@mail.ru

Abstract[1]

We propose a knowledge management-based approach to the information decision support of a manager charged with complex systems control tasks. The knowledge base of the decision support system consists of three knowledge representation models. These models are as follows: general knowledge model, Case Based Reasoning model and rules model. General knowledge model is reflected in domain ontology. Manager's experience is mapped to the CBR model. The model of rules reflects the expert knowledge contained in regulating documents. A knowledge base structure is proposed. An approach to the information decision support has been applied to the waste-water treatment facility management. The system has been implemented in accordance with Semantic Web specifications.

1. Introduction

During the recent years, rising complexity of problems related to wastewater treatment can be observed. Major issues arise from increasing requirements for water purification [5]. Even though measurement and control technologies are improving, the problem of incomplete or missing data still exists because many parameters are difficult to determine or cannot be determined at all. Furthermore, in specific cases, the measured data might not be representative for the overall system. Therefore, it often happens that a wastewater treatment plant (WWTP) manager must control the plant using strong theoretical background and experience gained from past events than with existing instructions. AI plays an important role in solving this problem.

In this paper, we present an approach to the developing of an information decision support system for a WWTP manager based on the integration of different AI techniques [2]. Theoretical background of the WWTP is proposed to be in a domain ontology. We also need an operator's experience in problem situations [5]. Even WWTP experts often are not able to determine the relevant influences caused the problem situation. Therefore, all information that may have significant impact on the problem is considered in the problem description. These knowledge will be stored in a proposed CBR-ontology.

The paper is structured as follows. In Section 2 we will present diverse models that reflect the proposed components of a DSS and model of a domain ontology. Section 3 will focus on developing a solution retrieval algorithm with specially defined similarity measures based on the ontological relations. Section 4 ends with conclusions.

2.Development of diverse knowledge presentation models for intellectual DSS.

Object-cognitive analyses (OCA) is used as the development methodology of the system [1]. OCA consists of object-oriented analysis (OOA), ontological analysis and the semantic web.

2.1. Modeling of knowledge presentation

We use OOA for domain analysis and modeling. In compliance with RUP (Rational Unified Process), the development of DSS comprises the following stages: business modeling, determinations of requirements, analysis and design, implementation. The following DSS modules are suggested (Figure 1): ontological knowledge base (OKB), solution retrieval engine and GUI. Modules implementation scenarios have been developed.

We use KAON2 [7] as a reasoner for our research. For reasoning, KAON2 supports the SHIQ(D), a subset of description logic. Knowledge management module is used for managing our ontologies. This process are being done in Protégé which is an ontology editor and knowledge-base framework [8].

Proceedings of the 8th International Workshop on Computer Science and Information Technologies CSIT'2006
Karlsruhe, Germany, 2006

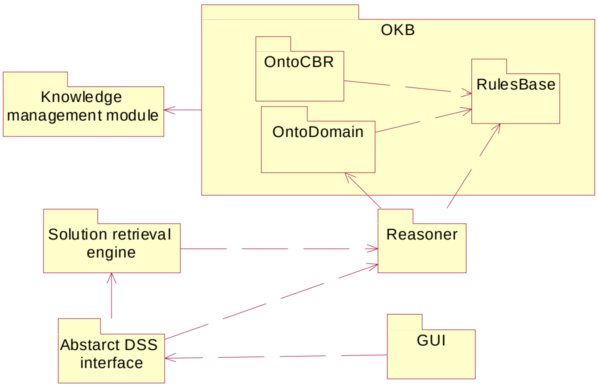

Figure 1: A model of DSS architecture

The novelty of the ontological knowledge base structure lies in the integration of common and special domain knowledge in the form of ontology (Figure 2). The common domain knowledge was extracted and formalized as the domain model of the ontology. The special domain knowledge were presented as cases distinguished by their object structure and capability to integrate with abstract domain knowledge.

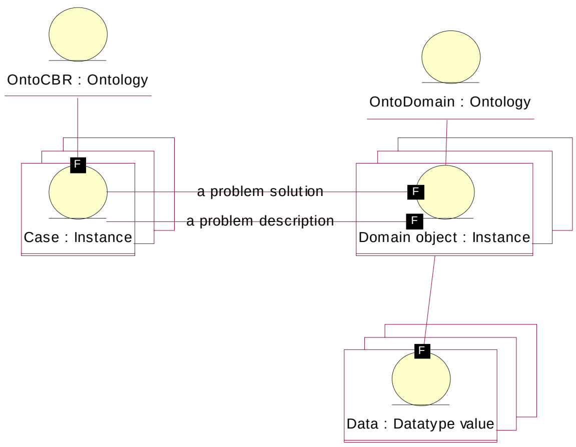

Figure 2: Diagram of OKB elements' collaboration

2.2.Development of a domain ontology model

The domain ontology is a part of the ontological knowledge base that formalizes the management context.
Ontology development techniques are task-specific due to their complexity. We propose staged ontology development based on the features of our particular domain and the analysis of existing techniques for different domains.

We propose knowledge acquisition and ontology conceptualization based on building of a semantic web including every possible relationships approved by experts. This step is mandatory due to the fact that the assessment of optionality and importance of relationships is a sufficiently complex task. We propose textual data of normative documents related to WWTP, glossary generated from UML models and expert knowledge, to be used as the input to building the semantic web of WWTP.

A conceptual model of the domain ontology was built in compliance with the proposed technique. The model allowed for building of domain objects taxonomy with abstract "Thing" as the root and the most common WWTP objects, such as "biological object", "water", "facility", "stuff" on the next level. Lower level describe objects specific to the particular treatment facility, such as "Active sludge", "Aerotank". The lowest level of the ontology is represented with individuals referenced from the CBR ontology.

3. Ontological knowledge base structure and retrieving algorithm

The following sections give a brief introduction to the Ontology Web Language. The reason for application of this language is explained in this section.

3.1. Choosing OWL for OKB formalization

The next step after the modelling of the ontological knowledge base is the development of appropriate formal models for it. The proposed ontological knowledge base (OKB) will be used by distributed operators of treatment facilities. That is why our OKB is based on the semantic web. The semantic web is used to make a wide range of web accessible data and services more readily accessible to automated processes. This is to be done by augmenting existing presentation markup with semantic markup. Ontologies play a key role in it. They are used as a source of shared and precisely defined terms that can be used in such metadata. The importance of ontologies in semantic markup has prompted the development of several ontology languages specifically designed for this purpose. We use Web Ontology Language (OWL) [6]. OWL is of particular significance as it has been developed by the W3C Web Ontology working group, and is now an official W3C recommendation. Our insight into the documents and capabilities available are based on keyword searches, abetted by clever use of document connectivity and usage patterns. OWL is intended to provide a language that can be used to describe the classes and relations between them that are inherent to web documents. We use OWL for formalizing the domain by defining classes and properties of those classes, defining individuals and assert properties about them, and reason about these classes and individuals to the degree permitted by the formal semantics of OWL.

We use OWL DL which is a sublanguage of OWL. OWL DL supports maximum expressiveness without losing computational completeness (all entailments are guaranteed to be computed) and decidability (all computations will finish in finite time) of reasoning systems. OWL DL includes all OWL language constructs with restrictions such as type separation (a class can not also be an individual or property, a property can not also be an individual or class). OWL DL is so named due to its correspondence with description logics, a field of research that has studied a particular decidable fragment of first order logic. OWL DL was designed to support the existing Description Logic business segment and has desirable computational properties for reasoning systems.

3.2. Formalizing OKB in OWL DL

We use OWL DL for describing OKB. The ontological knowledge base was developed in accordance with the meta-ontology. $Onto^{meta} = <C^{meta}, I^{meta}, R^{meta}, V^{meta}, Ax^{meta}>$, where $Onto^{meta}$ is a meta-ontology, which contains C^{meta}– classes, I^{meta} – instances, $R^{meta} = \{R,P\}$, R^{meta} – paradigmatic relations which up to date include causal relations, “is-a” and “part-of” relations, P – properties, V – properties values, Ax – axioms (rules). Thus, the meta-ontology formalizes such basic entities as instances, properties, rules/axioms and predefined relationships: aggregation, association and causality. The structure of OKB has been developed in accordance with UML diagram (Figure 1). $Onto^{KB} =< Onto^{domain}, Onto^{cbr} >$, where $Onto^{KB}$ is the ontological knowledge base, $Onto^{domain}$ is the domain ontology of the process of controlling wastewater treatment, $Onto^{CBR}$ is a specially structured ontology of problem situations cases.

A fragment of OWL DL $Onto^{domain}$ presentation is given below. A class “Active sludge” has been described.

```
=============Active_sludge===================
 <owl:Class rdf:about="#Active_sludge">
   <owl:disjointWith rdf:resource="#Situation"/>
   <owl:disjointWith>
     <owl:Class rdf:about="#Biological_object"/>
   </owl:disjointWith>
   <owl:disjointWith rdf:resource="#Biological_process"/>
   <rdfs:subClassOf>
     <owl:Restriction>
       <owl:onProperty>
         <owl:ObjectProperty
rdf:ID="containBiologicalObject"/>
       </owl:onProperty>
       <owl:allValuesFrom>
         <owl:Class rdf:about="#Biological_object"/>
       </owl:allValuesFrom>
     </owl:Restriction>
   </rdfs:subClassOf>
   <rdfs:subClassOf>
     <owl:Class rdf:about="#Object"/>
   </rdfs:subClassOf>
   <rdfs:label xml:lang="en">Active sludge</rdfs:label>
   <rdfs:comment xml:lang="ru">A major biological
component of WWT process
</rdfs:comment>
   <rdfs:subClassOf>
     <owl:Restriction>
       <owl:someValuesFrom>
         <owl:Class rdf:about="#Biological_object"/>
       </owl:someValuesFrom>
       <owl:onProperty>
         <owl:ObjectProperty
rdf:about="#containBiologicalObject"/>
       </owl:onProperty>
     </owl:Restriction>
   </rdfs:subClassOf>
 </owl:Class>
==========================================
```

The cases in $Onto^{CBR}$ have object-oriented structure. A case is an object expressed by the set of elements. This set consists of a problem description and a solution description. In the following, we give an overview of the structure of these two parts. The aim of the problem description is to characterize the current state of WWTP when a problem is observed. Even WWTP experts are often unable to determine all the factors contributing to a problem exactly. Therefore, all information that may have significant impact on the problem is considered in the description of a problem. Each case consists of objects (*I*), properties (*P*) and properties values (*V*). Each case is associated with a problem situation (an instance of a class “problem situation”).

i^{cbr}_i = ($category_i$, c_name_i, P_i ,Dec_i, Sc_i), where c_name_i – name of case i in $Onto^{cbr}$, $P_i = \{io_k, is_m\}$, i=1..m, m-set of descriptions in cases, $io_k = <P_i,V>$, P_i= <name, o_type>, type $\in C^{domain}$, name – property name, o_type – object property; $is_m = <P_j, V>$, где is_m - datatype property which is taken from $Onto^{domain}_i$, pr_i =<name, d_type>, d_type ={string, integer, boolean}, Dec_i – decision $I^{DEC}_i \in I^{domain}$ .Sc_i – an activity scenario, that has the same structure as a decision.

3.3. Developing a retrieval algorithm for finding solution in problem situations in OKB

The developed solutions retrieval algorithm contains logical deduction based on decision-making rules and description logic axioms. These rules and axioms have been developed by domain experts with the help of the conceptual ontology model and cluster analysis procedures. The input information for the procedures is extracted from laboratory measurements data.

The identification of a problem situation is based on the association of the parameters set A_i (describing the state of treatment facility in the particular moment) with a class of possible problem situations $cat_j \in Cat$.

$$A^n = \bigcup_{i=1}^{n} A_i, \quad A_i = (a_1, .., a_i, .., a_j, ..a_m), a_i \cap a_j = \varnothing, i \neq j, \psi : A_i \Rightarrow Category$$

A

$$\forall \left(A_i \in A^n \right) \exists \left(category_j \in Category : \Psi(A_i) = category \right)$$

control action d_j, associated with the resulting class can be recommended to the operator. Rules of identification based on the cluster analysis were developed by a domain expert and later formalized in SWRL (Semantic Web Rule Language, a Horn clause rules extension to OWL [3]). The structure of these rules is as follows:

Ax_n: $C_1(?x) \wedge C_2(?y) \wedge C_1.p_1.C_2(?x, ?y) \wedge C_3(?x, ?z) \rightarrow C_2(?z, ?y)$,
where $(C_1, C_2, C_3) \in C^{meta}$, $p_1 \in P^{meta}$, x,y are either variables or OWL individuals, and z is either a variable or an OWL data value.
The $Onto^{CBR}$ was developed in accordance with the rules mentioned above. Second part of the proposed algorithm is based on analogical retrieval of cases. A distinguishing feature of the algorithm is the use of proposed similarity measures for the ontological classes and properties [4]. A query q is created after the initialization of the retrieval procedure. This query includes a set of class instances in accordance with current WWTP parameters (A_i). Firstly, our system identifies the class of the problem situation. The query is then mapped to one of predefined classes with the help of rules. Then all the instances I_{c_q} of the class cat_i are retrieved and compared with the query with consideration of ontological relations.
After that, we determine, how much query elements and each instance of the determined class have in common. The "is-a" similarity considers the generalization relations and shows how many superclasses share the two classes C_q and C_i. To calculate the supercclasses similarity we define the set of class' superclasses as

$$C_q = \{C_q \in C : R(C_i, C_q) = "is-a" \vee C_q = C_i\}$$

$$Sim_{isa}(C_q, C_i) = \frac{|C_q \cap C_i|}{|C_q \cup C_i|}$$

where C_q –classes of the query, C_i- classes of the $Onto^{doamin}$, R – «is-a»relation between classes.

Second, as classes have properties, that are instances of other classes (object properties) or instances of primitive types (datatypes) related with a "part-of" relation. We have developed a special similarity measure ($Sim_{partof}(q_i, I_n)$) to compare them in order.

Since attributes are instances of different types, several attributes similarity calculation functions should be applied. For the comparison of object properties (in case the property is an instance of another class), we apply similarity measure formula S_{glob} recursively. In that case, we limit the depth of the recursion with some threshold, so that infinite loop condition wouldn't occur. Separate functions are provided for the primitive types, such as string (Hamming distance), numbers (distance for numbers is: $fsim_T(a,b) = 1 - \frac{|a-b|)}{ceil_K}$, where $a, b \in K$, $ceil_K$ is a maximum value in K) and Boolean. We apply weights [10], assigned to the attributes here.

The "part-of" similarity is calculated as follows:

$$Sim_{partof}(q_i, I_n) = \frac{\sum_{j=1}^{l}\sum_{k=1}^{m} fsim_t(rq^i_j, ri^{i'}_k) * w_j}{l+m}, t \in T, r \in R$$

where $rq^i_1 \dots rq^i_k$, $ri^i_1 \dots ri^i_k$ – properties of matching type and name in both instances being compared, m – the sum of the number of all attributes bound to each instance, fsimt is a function used for similarity calculation of properties belonging to type t, T = {"string", "integer", "boolean", } or S_{glob}, w_k – weight, assigned to the attribute indexed k.
Having the local similarity measures for the query and the case defined, we can calculate the global similarity measure as follows:

$$S_{glob}(q,i) = \frac{Sim_{isa}(q,i) * w_{isa} + Sim_{partof}(q,i) * w_{partof}}{2}$$

4 Conclusion

In the scope of our work we defined an approach to the information decision support in problem situations of a WWTP manager. We proposed a structure of an ontological knowledge base, which integrates general knowledge about WWTP and knowledge about concrete problem situations. Such a structure would help to find a solution in problem situation because of the more complete domain description and consideration of ontological relations in the developed algorithm. The algorithm was developed in accordance with Semantic Web architecture that allows to consolidate distributed knowledge and are used by distributed operators of treatment facilities. Such an architecture makes grouping information of different origin into a single entity possible. Furthermore, the proposed similarity measure makes it possible to consider different types of properties and relations. This allows for comparison of information structures known at the design time as well as for seamless extension by adding new types and defining new relations. The technical implementation of the research is being developed.

References

1. Badamshin R.A., Ilyasov B.G., Cherniakhovskaya L.R. "The problems of the complex dynamic objects control in critical situations on the base of knowledge". Mashinostroenie, Moscow, 2003.
2. Bergmann, Ralph, Schaaf, Martin. "*On the Relations between Structural Case-Based Reasoning and Ontology-based Knowledge Management*", The Journal of Universal Computer Science 9 (7), pages 608-626, 2003.
3. Horrocks I., Patel-Schneider P. F., Bechhofer S. and Tsarkov D. *OWL rules: A proposal and prototype implementation* // Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 3(1), pages 23-40, 2005.
4. Shkundina R., Schwarz, S.: "*A Similarity Measure for Task Contexts* "// Proceedings of the 6th International Conference on Case-Based Reasoning (ICCBR) 2005, LNCS 3620, Springer Verlag, 2005. pages 261-270, Chicago, 2005.
5. Wiese, Jürgen, Stahl, Armin, Hansen, Joachim: "*Applying and optimizing case-based reasoning for wastewater treatment systems*", AI Communications.

Special Issue: Binding Environmental Sciences and AI 18(4), pages 269-279, 2005.
6. Web Ontology Language (OWL)//The World Wide Web Consortium // http://www.w3.org/2004/OWL
7. KAON2. http://kaon2.semanticweb.org
8. Protégé. http://protege.stanford.edu/