UDK 004.75

Artem Kaplunov, Bohdan Ivanishchev, Oleksandr Honcharenko.

USING ONTOLOGICAL EXTENSIONS WHILE TASKS MONITORING IN A DISTRIBUTED COMPUTING SYSTEM

The article analyzes the feasibility of using ontologies and the method of information addition to increase the efficiency of information systems.

Key words: ontology, information theory, distributed system.

Tabl.: 2.

The relevance of those studies. Taking into account complication of the technical and social infrastructure, information received acquired status of the new strategic resource.

The acquisition by information of the role of a strategic resource contributed to the development of new information technologies for the receipt, processing, storage and provision of large volumes of information to users. In this way, any attempts to increase the efficiency of such systems will have significant positive consequences for all areas of our lives.

Problem formulation. The most common way in the literature to measure the quality of information systems (IS) is to use the "recall" and "precision" evaluation characteristics. "Recall" is the ratio of the number of relevant documents found to the number of stored. "Precision" displays the ratio of the number of relevant documents found by the IS on an arbitrary request to the number of all found. However, it is advisable to apply these criteria only in IS that operate with poorly structured and informal information.

In the information systems considered in the work, the "recall" and "precision" are 100%, and the use of these characteristics to assess the quality of their work is meaningless. Usually assessment of the effectiveness of such information systems is carried out taking into account their operational characteristics. Particularly the evaluation criteria are the query processing time, the amount of memory used, the degree of processor(s) load, etc. Although these are important characteristics of an IS, they do not allow one to determine the real effectiveness of information storage and processing.

Analysis of recent scientific researches and issues. Nowadays, when designing and developing information systems, the use of semantic technologies, such as ontologies, has become mandatory. Ontologies are the formalization of a certain field of knowledge in the form of a multitude of concepts and relations between them. Ontologies are widely used in computer science as a means to represent knowledge and process data taking into account their semantics.

The ontology language defines the syntax and semantics of the ontology in such a way that it can be unambiguously perceived and interpreted by the computer. One of the leading languages of ontologies today is considered the OWL language, standardized by the World Wide Web Consortium. In its turn descriptive logic (DL) [1] provides a well-studied logical basis for ontological languages. Today, there are many diverse DL languages, differing in the degree of expressiveness and computational complexity.

Oracle added ontology support to its Oracle 11g object-relational database management system [2]. However, this DBMS provides a projection of RDF-models entered as data and applies a lot of generating rules to it in accordance with the RDF(S) and OWL2 RL specifications. In this case, rather weak if-semantics are used, expressed as IF ... THEN rules. Data queries are made through SQL in SPARQL queries or using the special ONT_RELATED and ONT_EXPAND statements. The queries used in Oracle 11g are extremely complex and cumbersome, and the weakened semantics and RDF-oriented model used are a significant drawback.

Specification of uninvestigated parts of general matters. An alternative approach, which considers the possibility of translating data from a relational database into its ontological representation, was proposed in [3]. This made possible the integration of RDBMS systems into the canvas of the Semantic Web. Unfortunately, the approach does not consider the analysis of the transmitted data taking into account their semantics. It also does not provide ontological additions or subsequent simplification of requests to the provided information.

A similar approach was proposed in [4], but in contrast to the previous work, it was proposed to use ontologies for access to the "underlying" IS. At the same time, a new special DL descriptive logic, Lite_A, and ontology-based SQL query generation mechanisms are considered. However, the method proposed by the authors is not universal and does not allow achieving the set goals.

Problem Statement. The purpose of this work is to substantiate the feasibility of using ontology and the method of information Supplement to improve the efficiency of information retrieval and storage by information systems. The approach proposed in the study can play a significant role in expanding the capabilities of information systems and improving their operational characteristics.

Statement of the main materials. Consider the concept of augmented IS S^+ as

$$S^+ = \langle X, A, V, \rho, \omega \rangle \ , \tag{1.1}$$

where X – set of IS objects, A – set of attributes, $V = \cup_{\alpha \in A} V_\alpha$ where V_α – set of attribute values $\alpha \in A$, ρ – definition function $\rho: X \times A \to V$ and ω – add-on function $\omega: X \times A \to V$.

The purpose of the function ω is to additionally determine the attributes of the object x, based on the set of already defined values specified by the function ρ.

We introduce the concept of ϱ_x^S as an aggregate description of an object x in an IS S (or ϱ_x if the IS S is defined by context). In this way ϱ_x^S for any object $x \in X$ will include the values of all attributes of this object:

$$\varrho_x^s = \cup_{\alpha \in A} \rho(x, \alpha) \tag{1.2}$$

Thus, the aggregate description of object $x \in X$ in the supplemented IS S^+ will be equal to:

$$\varrho_x^s = \bigcup_{\alpha \in A} \rho(x, \alpha) \cup \omega(x, \alpha) \tag{1.3}$$

Many methods can be used to implement the function *ω*. An ontological approach will be considered for these purposes in our work.

Thus, we consider the ontologically supplemented IS S_O^+ as an extension of the definition (1.3):

$$S_O^+ = \langle X, A, V, \rho, \omega, O \rangle, \tag{1.4}$$

where O is the ontology of the IS.

In order to quantify the amount of information that an arbitrary IS has, we turn to the theory of information [5].

For a random variable *X*, having a finite number of values, so $P(x_i) = \rho_i, \rho_i \geq 0, i = 1,2, \dots, n$ and $\sum_{i=1}^{n} \rho_i = 1$ for calculating the entropy according to the Shannon formula:

$$H = -\sum_{i=0}^{n} \rho_i log_2 \rho_i = \sum_{i=0}^{n} \rho_i log_2 \frac{1}{\rho_i} \tag{1.5}$$

In the particular case, when all options are equally probable, there remains a dependence only on the number of options considered and the amount of information *I* in this case is determined as the binary logarithm of the number of states:

$$I = log_2 n \tag{1.6}$$

In addition, we introduce the concept of the aggregate stored information γ^S of the IS S, which will determine the set of all statements in this IS:

$$\gamma^S = \cup_{x \in X} \varrho_x^s \tag{1.7}$$

Based on the above principles, we will try to estimate the amount of information stored in the IS S:

$$I(S) = I\left(\gamma^S\right),$$

$$I\left(\gamma^S\right) = \sum_{x \in X} I(\varrho_x^s),$$

$$I(\varrho_x^s) = \sum_{\alpha \in A} log_2 |V_\alpha|, \text{ where } \forall \rho(x, \alpha) \neq \emptyset. \tag{1.8}$$

Thus, in an ontologically supplemented IS, knowledge presented in the form of an ontology allows one to perform a logical analysis on information objects $x \in X$ and to further define their attributes, thereby increasing the amount of information in the IS. That is, for all cases the statement $I\left(\gamma^{S_O^+}\right) \geq I\left(\gamma^S\right)$ will be true.

Consider an example of an ontologically supplemented IS S_O^+ for monitoring tasks in a distributed computing system:

$$X = \{x_1, x_2, x_3, x_4, x_5\},$$

$$A = \begin{Bmatrix} runningtasks, status \\ IPaddress, tasks\text{in}thequeue \end{Bmatrix},$$

$$V = \begin{Bmatrix} V_{IPaddress} \cup V_{tasks\in thequeue} \cup \\ V_{runningtasks} \cup V_{status} \end{Bmatrix}$$

$$V_{IPaddress} = \{x \in N | 0 \leq x \leq 2^{32} - 1\},$$

$$V_{tasks\in thequeue} = N, V_{runningtasks} = N,$$

$$V_{status} = \{idle, busy, inaccessible\}.$$

The function ρ is expressed in the form of table 1:

Table 1

Function ρ

X	IP address	tasks in line	running tasks	status
x_1	192.168.0.1	24	2	
x_2	192.168.0.2	1	1	
x_3	192.168.0.3	3	0	
x_4	192.168.10.4	1		busy
x_5	192.168.10.5			idle

Suppose the IS was supplemented by the following ontology:

$$O = \begin{cases} \exists status.idle \equiv \exists tasks\text{in}thequeue.(=, 0) \\ \sqcap \exists runningtasks.(=, 0) \\ \exists status.busy \equiv \exists runningtasks.(\geq, 1) \\ \exists status.unavailable \equiv \exists tasks \in thequeue.(>, 0) \\ \sqcap \exists startedtasks.(=, 0) \end{cases}$$

In this case, the predefined IS will have the following form (Table 2):

Table 2

Function ρ+ω

X	IP address	tasks in line	running tasks	status
x_1	192.168.0.1	24	2	busy
x_2	192.168.0.2	1	1	busy
x_3	192.168.0.3	3	0	inaccessible
x_4	192.168.10.4	1	≥1	busy
x_5	192.168.10.5	0	0	idle

Using the expressions (1.3) and (1.8) it is easy to calculate that initially the IS contained about 387 bits of information. The field of ontological addition is the amount of information in the IS amounted to 456 bits. At the same time, for the object x_4, the uncertainty (entropy) with respect to the "running tasks" parameter was reduced.

Now let's look at how ontologies can significantly reduce the complexity of a query to an IS.

Let us turn to the ontologically supplemented IS from the example. We will make a search query to the IS, which will return a list of all idle or soon available computing resources in the local network:

$$Q_S = (tasks \in queue: 0) \wedge (runningtasks: \leq 1) \wedge (IPaddress: \geq 192.168.0.1) \wedge (IPaddress: \leq 192.168.0.255)$$

We expand the ontology of the IS S_O^+ with the following expressions:

$$O = \begin{cases} \exists IPaddress.(\geq 192.168.0.1, \leq 192.168.0.255) \subseteq Localresource \\ \exists status.idle \cup \exists runningtask.(=, 1) \\ \sqcap \exists tasks \in queue.(=, 0) \subseteq Unloaded \end{cases}$$

At the same time, two new terms will be added to the set of terms T_S of the query language L_S from the ontology: Local resource and Unloaded. In view of the new terms, we rewrite the search query as follows:

$$Q_{S_O^+} = LocalResource \wedge Unloaded$$

Using the formula (1.4), we can calculate that the amount of information in the initial request was 128 bits. In the second case, the amount of information was only 2 bits, since the terms defined in the ontology do not have a range of permissible values and only express a certain concept. Naturally, a request to an IS can and, most likely, will contain both terms defined in the ontology and ordinary terms based on descriptors.

Conclusion. We were able to significantly simplify the request for IS, while maintaining accuracy and completeness.

There may be a misconception that ontology has increased the complexity of the IS, since it will take much longer to create it than to compile a complex query. However, ontologies are compiled to describe an entire area of knowledge, providing a wide range of frequently used terms for reuse.

References

1. Baader F. The Description Logic Handbook: Theory, Implementation, and Applications. (2ed) / F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. PatelSchneider. – Cambridge University Press. – 2007.

2. Murray C. Oracle Database Semantic Technologies Developer's Guide, 11g Release 2 (11.2), E25609-03 // Oracle. – 2012. – Режим доступу до ресурсу: http://docs.oracle.com/cd/E11882_01/appdev.112/ e25609/toc.htm.

3. Sedighi S. Semantic Query in a Relational Database Using a Local Ontology Construction / S. Sedighi, R. Javidan // South African Journal of Science. – 2012. – 108(11/12). – C. 1-10.

4. Poggi A. Linking Data to Ontologies / A. Poggi, D. Lembo, D. Calvanese, G. Giacomo, M. Lenzerini, R. Rosati // Journal on data semantics. – Springer, – 2012. – Vol X. – C. 133-173.

5. Shannon C.E. The Mathematical Theory of Communication. / C. Shannon, W. Weaver, R. Blahut. – Urbana: University of Illinois press, – 1949.

AUTHORS

Artem Kaplunov – PhD student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: art.kaplunov@gmail.com

Bohdan Ivanishchev – PhD student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: art.kaplunov@gmail.com

Oleksandr Honcharenko – student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”

E-mail: alexandr.ik97@ukr.net