Development of a technology for collecting and analyzing data for monitoring based on an ontological approach

Yuri I. Molorodov^{1,2}, Oleg V. Kasatkin²

¹Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia ²Novosibirsk State University, Novosibirsk, Russia

Abstract

One of the ways to build information models is ontological modeling. The use of ontologies greatly facilitates the exchange of data between embedded models and utilities for the digital representation of an object or a real-world system, sometimes called "digital twin" (DT). It is also important to establish a correspondence between the DT, people and external programs. Based on the dictionary of the main terms, classes, objects of the subject area and the relations between them, we have built an ontology of the hydroelectric dam DT.

Keywords

Data collection and analysis system, monitoring of the technical condition of hydroelectric dams, dynamic characteristics, processing of seismometric monitoring data, digital twin ontology of hydroelectric dams.

1. Introduction

A digital twin is a software analogue of a physical device that simulates internal processes, technical characteristics and behavior of a real object under environmental influences. The concept of digital twins implies the connection of the physical and digital world through the interaction of information models. In other words, a mathematical model is created for a physical object, a piece of equipment or an entire process, which is then used to analyze the behavior of the object. Moreover, the digital model is constantly updated to fully correspond to the current state of the real object. This makes it possible to identify unexpected changes in processes, optimize the operating modes of equipment, prevent breakdowns and accidents, which ultimately allows you to significantly increase the reliability and efficiency of operation.

According to regulatory documents, hydroelectric power plants are designed with a minimum expectation for 100 years of operation, but with proper maintenance, the actual operation period can be significantly increased and reach several hundred years [8]. The main equipment and structures naturally wear out and require major repairs, modernization, replacement. The entire life cycle of a hydroelectric power plant is a series of constant updates, upgrades, reconstructions associated with the replacement of automation systems or with the replacement of main or auxiliary equipment.

SDM-2021: All-Russian conference, August 24-27, 2021, Novosibirsk, Russia

[🛆] yumo@ict.sbras.ru (Y. I. Molorodov); o.kasatkin@g.nsu.ru (O. V. Kasatkin)

^{© 0 2021} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

The use of a digital twin of the hydroelectric dam will significantly increase the efficiency of operation, as well as determine the most accurate time of maintenance.

2. The use of digital twin

The digital twin uses information from sensors (sensors) installed on a real object not only to update the current digital copy, but also to monitor the state and analyze the residual resource of the object, predict defects.

To build a digital twin of the dam, first of all, it is necessary to build an information model. One of the ways to build information models is ontological modeling. The use of ontologies greatly facilitates the exchange of data between embedded models and utility programs of the digital double, as well as between the digital double and people or external programs.

There are many definitions of ontology of varying degrees of formalization. Often, the definition is given based on the methods of constructing an ontology. Nevertheless, despite the abundance of approaches to the definition, the same composition of components is usually used, which includes:

- classes or concepts that are usually interpreted quite broadly and can include entities of any kind;
- 2) instances, i.e. separate entities, the totality of which forms classes;
- 3) attributes specific internal properties of classes and instances;
- relationships mean the relationship between the classes of the subject area, for example, the relation of taxonomy;
- 5) axioms, or rules of inference, which determine some always true statements that serve, for example, to check information for correctness.



Figure 1: Ontological description of the digital twin of the dam.

According to [6], the ontology is formally defined as follows:

$$O_R = \langle C, R, A, T, D \rangle,$$

where $C = \{C_1, \ldots, C_n\}$ – set of domain classes; $R = \{R_1, \ldots, R_m\}$, $R_i \subseteq C \times C$ – the set of relations defined on the classes of the subject area; $A = \{a_1, \ldots, a_w\}$ – a set of attributes describing the properties of concepts C and relations R; T is the set of standard attribute values; $D = \{d_1, \ldots, d_n\}$ – the set of domains.

To build an ontology of the digital twin of a hydroelectric dam, we will use the CmapTools program. Based on the works [2, 3, 12], as well as following the instructions of the methodological manual [5], the ontologies presented in Figure 1 were constructed.

Figure 1 shows that the model of a digital twin of a hydroelectric dam requires constant updating of simulation models and accumulation of data obtained from a real object to monitor the current technical condition and predict the remaining resource. To solve these problems, an information system for collecting and processing information is needed, which, along with calculation and information models, can become a central link in the construction of a digital twin of the dam.

3. Functionality and structure of the system

In the work [1], a method for assessing the technical condition of buildings and structures is proposed, based on the analysis of changes in dynamic characteristics determined under the influence of a microseismic background of a natural and man-made nature, i.e. without the use of special excitation sources. These characteristics are manifested under dynamic loads and reflect the technical condition of the building structure as a whole (frequencies and forms of natural vibrations, attenuation decrements, statistical characteristics, etc.). The analysis of trends in dynamic characteristics, taking into account the influence of seasonal environmental factors, allows us to determine their irreversible changes and monitor the current technical condition of the entire structure.

To perform the tasks of monitoring the state of the object in this work, the periodic (according to a given schedule) registration of dam vibrations under the influence of microseismic background and dynamic loads from the equipment operating at the HPP is selected.

Based on the analysis of regulatory documents, the following main functional requirements for the system were identified:

- 1) the possibility of scheduled and periodic registration of microseismic vibrations of the dam;
- the system must store records of seismometric monitoring data and information about the functioning of the system;
- the possibility of data analysis using various information processing algorithms, calculation of the studied parameters and characteristics;
- 4) the ability to log error messages in the system, notify the user in case of exceeding the permissible values of the observed values and problems in the system;
- 5) visualization of the analysis results.

Figure 2: System functionality.



Figure 3: The basic structure of the system for collecting, storing, analyzing and visualizing data arrays.

In general, the operating procedure and functionality of the system under development are shown in Figure 2.

The assessment of the technical condition of hydraulic structures is carried out in accordance with the criteria adopted in the current building codes and rules for ensuring the safety of operation of buildings and structures according to the theory of limit states. In the developed system, the criteria for the safety of hydraulic structures in terms of seismometric monitoring should be established in the form of functional dependencies between the dynamic characteristics and parameters of external influences and loads. These dependencies are determined based on the results of statistical analysis of data at the accumulation stage.

The structurally developed system should be a construction of related subsystems: data collection, storage of the received data arrays, information processing and analysis, and data visualization. Figure 3 shows the basic structure of the system for collecting, storing, analyzing and visualizing data arrays.

The decomposition of the system into functional modules with well-defined data flows between them can significantly reduce the complexity of the system and its development, as well as increase the versatility of the modules.

3.1. Development of a data processing and analysis module

The data processing and analysis module was written in the C++ programming language in the Microsoft Visual Studio 2019 development environment using the STL library.

3.2. Input data format

As an input in this paper, we will consider the CIBF seismometric monitoring file format used in the operation of the automated earthquake registration system and monitoring of the technical condition of the dam (PAK-MZ) of the Krasnoyarsk HPP. This system periodically registers stationary micro-vibrations of the dam at ten observation points in three directions of vibration of the structure: longitudinal, transverse and vertical.

The CIBF file format is a specially developed format for seismometric monitoring data. The file name includes the start time of data registration (the time of the first count) in the format dd mm yyy hh mm ss.cibf. All data is stored in the little-endian format (i.e. the lowest byte comes first).

CIBF files consist of data packets (in one packet, data for all channels). Each packet contains a header and data. The header includes the serial number of the packet for monitoring data transmission, the parameters of the registration profile and the length of the data. Each packet contains a 32-byte header and the data itself. The header structure is shown in Table 1.

The number of samples (the Sample Received field) shows how many samples are in the data following the header. Thus, the headers can be used for iterating over the packages of the file.

The data is presented as a set of samples by channels. Each sample is represented as two bytes of the data itself (the code with the ADC) and two bytes of service information (Table 2).

The bit fields F1, F2, and F3 are responsible for channel error, negative saturation, and positive saturation, respectively. If the 8th bit of F1 is 0, then a channel error has occurred. If the 9th bit of F2 is not equal to 0, then negative saturation has occurred. If the $10^{\rm th}$ bid of F3 is not equal to 0, then a positive saturation has occurred. The lowest byte represents a cyclic code from the range from 0 to 254.

Byte number	Field name	Туре	Length	Comment
0	Command	word	2	1: test, 2: data, 3: service
2	SamplesResived	word	2	Number of samples
4	DateTime	TDateTime (double)	8	Start time
12	PackedSize	integer	4	The size of the data packet after the header
16	PackedNmb	integer	4	Package number
20	SendDateTime	TDateTime (double)	8	The time of sending – possibly an empty field, a reserve
28	Reserved0	word	2	0
30	Reserved1	word	2	0

Table 1 Package Header structure.

Table 2

Format of service information ctrl.

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Ch	а	nn	e	1	F3	F2	F1	Су	С	li	С		С	0	de

3.3. Data extraction and processing

To read CIBF format files, we will use the functionality of the fstream header file. Let's create an object of the std::ifstream class and link the class object to the CIBF file from which the reading will be performed (hereinafter we will simply call it "file"). At the same time, we need to receive data in binary form, so we will use the binary mode of opening the file. If it is impossible to open the specified file, we will generate an exception of the std::exception type.

Recall that the smallest addressable element of the computer's RAM is a byte. Let's calculate the exact length of the file. To do this, use the seekg function to move the file pointer to the final position and use the tellg function to find out its number (in bytes), then use the seekg function again to return the pointer to the starting position for further work.

Note that the computer's central processor uses a machine word as the main unit of work with memory, the size of which depends on the processor architecture. When storing an arbitrary object in memory, it may happen that some field consisting of several bytes will cross the "natural boundary" of machine words in memory. Some processor models cannot access such data in memory, or they take longer to access data located inside an entire machine word in memory. Therefore, some C++ compilers can use automatic "data alignment". This optimization consists in inserting insignificant bytes between the fields of the data structure when placing it in memory to speed up access to this structure.

For the convenience of reading, we will describe data packages using structures — the custom data type struct of the C++ language. Each packet will be a structure, whose members will be other structures — the packet header and an array of samples. When describing structures, we will use the pragma pack(push,1) and pragma pack(pop) preprocessor directives to exclude any data alignment.

Create an integer variable with a zero initial value to track the number of the current position (in bytes) from the beginning of the file. Further, everywhere under the word "reading" we will mean byte-by-byte extraction of the structure into a variable of the corresponding type using the read operator. The process of reading data packets will occur in the body of the main while loop, the condition for the completion of which will be reaching or exceeding the number of the current position of the file length value. The increment of the value of the current position number will occur after using the read operator.

In the body of the main loop, we will first read the packet header. At the beginning of the header of each package is a number that characterizes the contents of the package. We will use it to control the reading. If the number 2 is the first in the read packet, then we determine how many samples are in the packet and read them into the buffer. Otherwise, we notify the user that this package contains service or test information.

In the cycle, we will process the received data according to the number of read samples. First of all, from the ctrl field, we will determine the number of the channel on which the countdown was received. Next, it is necessary to convert the value N obtained from the ADC to the form

$$A = \frac{20N}{2^{16}} \cdot \frac{1}{K} \cdot \frac{1}{KEMS},$$

where A is the acceleration amplitude; K is the total signal gain; KEMS is the electromechanical coupling coefficient. The KEMS of the sensor determines its sensitivity. At the

Tab	le	3	

Pavilion, mark, section, component	Gain
1k PV.1, mark.223, sec.Ba, Z	$100 \times 4 = 400$
2k PV.1, mark.223, sec.Ba, Y	$100 \times 4 = 400$
3k PV.1, mark.223, sec.Ba, X	$100 \times 4 = 400$
4k PV.2, mark.223, sec.8, Z	$100 \times 2 = 200$
5k PV.2, mark.223, sec.8, Y	$100 \times 2 = 200$
6k PV.2, mark.223, sec.8, X	$100 \times 1 = 100$
7k PV.3, mark.132, sec.8, Z	$100 \times 4 = 400$
8k PV.3, mark.132, sec.8, Y	$100 \times 4 = 400$
9k PV.3, mark.132, sec.8, X	$100 \times 4 = 400$
10k PV.4, mark.223, sec.22, Z	$100 \times 1 = 100$
11k PV.4, mark.223, sec.22, Y	$100 \times 1 = 100$
12k PV.4, mark.223, sec.22, X	$100 \times 1 = 100$
13k PV.5, mark.132, sec.22, Z	$100 \times 4 = 400$
14k PV.5, mark.132, sec.22, Y	$100 \times 4 = 400$
15k PV.5, mark.132, sec.22, X	$100 \times 4 = 400$
16k	Reserved
17k PV.6, mark.244, sec.37, Z	$100 \times 1 = 100$
18k PV.6, mark.244, sec.37, Y	$100 \times 1 = 100$
19k PV.6, mark.244, sec.37, X	$10 \times 4 = 40$
20k PV.7, mark.132, sec.37, Z	$100 \times 2 = 200$
21k PV.7, mark.132, sec.37, Y	$100 \times 2 = 200$
22k PV.7, mark.132, sec.37, X	$100 \times 2 = 200$
23k PV.8, mark.223, sec.54, Z	$100 \times 1 = 100$
24k PV.8, mark.223, sec.54, Y	$100 \times 1 = 100$
25k PV.8, mark.223, sec.54, X	$10 \times 4 = 40$
26k PV.9, mark.132, sec.54, Z	$100 \times 2 = 2000$
27k PV.9, mark.132, sec.54, Y	$100 \times 2 = 200$
28k PV.9, mark.132, sec.54, X	$100 \times 2 = 200$
29k PV.10, mark.223, sec.71, Z	$100 \times 4 = 400$
30k PV.10, mark.223, sec.71, Y	$100 \times 4 = 400$
31k PV.10, mark.223, sec.71, X	$100 \times 4 = 400$
32k	Reserved

Krasnoyarsk HPP dam, $KEMS = 2V \sec^2/m$. The total signal gain is calculated as the product of the pre-gain coefficients K1 and the main gain K2. The values of the gain coefficients K are shown in Table 3.

The obtained acceleration amplitudes are distributed over channels using objects of the std container class::vector.

3.4. Presentation and analysis of the results of the module

The implemented module was tested on CIBF files. The correctness of the module operation and the results obtained was confirmed by a specialist of the PAK-MZ system of the Krasnoyarsk

HES.

Based on the obtained values of the acceleration amplitudes, the statistical characteristics of the oscillations are calculated. An example of the calculated statistical characteristics for one of the monitoring sessions is shown in Table 4.

One of the methods for visualizing the process of changing the amplitudes of accelerations over time is the construction of seismic tracks. A seismic trace is a graph of the dependence of the signal level of seismic waves or noise on the time of their registration.

Let's use the Gnuplot charting program. To build a seismic track, it is necessary to know the duration of the registration of vibrations (512 seconds) and the number of recorded samples (the length of the object std::vector, where the data is stored after extraction). If we set aside the time in seconds horizontally, and the values of the acceleration amplitudes (mm/s^2) vertically,

Tał	ole	4
-----	-----	---

Channel No.	Coordinate	Average, mm/s^2	Variance, mm^2/s^4	Frequency, Hz
1	Z	-0.02	0.01	4.63
2	Y	-0.01	0.01	4.87
3	Х	-0.02	0.04	4.33
4	Z	-0.04	0.16	4.73
5	Y	-0.03	0.09	5.81
6	Х	-0.08	4.84	4.32
7	Z	-0.02	0.01	4.5
8	Y	-0.01	0.01	4.89
9	Х	-0.02	0.04	4.12
10	Z	-0.09	0.11	4.71
11	Y	-0.07	0.1	5.2
12	Х	-0.06	0.63	4.31
13	Z	-0.02	0.04	4.53
14	Y	-0.02	0.02	4.03
15	Х	-0.02	0.14	4.36
17	Z	-0.08	0.12	4.72
18	Y	-0.06	0.13	6.02
19	Х	-0.12	3.9	4.38
20	Z	-0.04	0.09	4.22
21	Y	-0.03	0.07	5.11
22	Х	-0.03	0.12	4.6
23	Z	-0.09	0.12	6.17
24	Y	-0.09	0.13	4.97
25	Х	-0.2	3.24	4.53
26	Z	-0.03	0.05	4.81
27	Y	0.07	0.03	5.54
28	Х	-0.05	0.1	4.34
29	Z	-0.02	0.01	7.31
30	Y	-0.03	0.02	5.98
31	Х	-0.02	0.04	4.87

Statistics for the session 03/04/2011/19:00.



Figure 4: Seismic route of the 2nd channel of the session 03/04/2011/07:00.

then we will get a visualization of the dam oscillation process in the specified observation point and direction in time on the graph. An example of a seismic route is shown in Figure 4.

4. Conclusion

On the basis of regulatory documents regulating the maintenance of hydroelectric dams, the subject area is investigated and an information system for collecting and analyzing seismometric monitoring data is developed. The basis for designing an information system is an ontological approach. Using this approach, the ontology of the digital twin of the hydroelectric dam is constructed. The main functional capabilities of the information system are formulated and its structure is described. The data processing and analysis module is implemented in the C++ programming language. The implemented module has been tested on CIBF-format seismometric monitoring data packages.

References

- Baryshev V.G., Kuzmenko A.P., Saburov V.S. et al. Dynamic test inspection of dams under the influence of operational dynamic loads // Gidrotehn. Construction. 2002. N. 10. P. 26–36. (In Russ.)
- [2] Borovkov A.I., Ryabov Yu.A. Digital twins: Definition, approaches and methods development / A.V. Babkin (Ed.) // Digital Transformation of Economy and Industry: Proceedings of the Scientific-Practical Conference. June 20–22, 2019. SPb.: Polytechnic Press, 2019. P. 234–245. (In Russ.)
- [3] Borovkov A.I., Ryabov Yu.A., Maruseva V.M. A new paradigm of digital design and modeling of globally competitive products of a new generation // Digital Production: Methods, Ecosystems, Technologies. MSU Skolkovo, 2018. P. 24–44. (In Russ.)
- [4] Bulletin of RusHydro-Science and Technology. The future digital everyday life of RusHydro. Available at: https://vestnik-rushydro.ru/articles/2-fevral-2019/nauka-i-tekhnologii/ budushchie-tsifrovye-budni-rusgidro (In Russ.)
- [5] Muromtsev D.I. Conceptual modeling of knowledge in the Concept Map system. St. Petersburg: SPb GU ITMO, 2009. 83 p. (In Russ.)

- [6] Zagorulko Yu.A., Sidorova E.A., Borovikova O.I. Ontological approach to the construction of information support systems for scientific and industrial activities // Materials of the All-Russian Conference with International Participation "Knowledge – Ontology – Theory" (ZONT-09). Novosibirsk, 2009. Vol. 2. P. 93–102. (In Russ.)
- [7] Korolenko D.B., Kuzmenko A.P., Moskvichev V.V., Saburov V.S. Information system of seismometric monitoring of the technical condition of hydraulic structures: Experience of modeling, development and implementation // Computational Technologies. 2019. Vol. 24. No. 5. P. 13–37. DOI:10.25743/ICT.2019.24.5.003. (In Russ.)
- [8] How many years can hydroelectric power plants last? Available at: http://blog.rushydro. ru/?p=9950 (accessed April 06, 2021). (In Russ.)
- [9] STO 70238424.27.140.003-2010. Hydrotechnical structures of hydroelectric power plants and hydroelectric power plants. Organization of operation and maintenance. Norms and requirements. Introduced 2010-09-30. Moscow: NP "Invel", 2010. 15 p. (In Russ.)
- [10] STO 70238424.27.140.035-2009. Hydroelectric power stations. Monitoring and evaluation of technical structures during operation. Norms and requirements. Introduced 2009-12-31. Moscow: NP "Invel", 2009. 59 p. (In Russ.)
- [11] STO RusHydro 02.01.80-2012. Hydrotechnical structures of hydroelectric power plants and hydroelectric power plants. Operating rules. Norms and requirements. Entered 2012-10-29. Moscow: JSC "RusHydro", 2012. 181 p. (In Russ.)
- [12] Parrott A., Warshaw L. Industry 4.0 and the digital twin technology. Deloitte Insigts (12-05-2017) — Manufacturing meets its match. Available at: https://www2.deloitte.com/us/ en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html.