

# The Approach to Applications Integration for World Data Center Interdisciplinary Scientific Investigations

Grzegorz Nowakowski

Department of Automatic Control and Information Technology, Faculty of Electrical and Computer Engineering, Cracow University of Technology Cracow, Poland  
gnowakowski@pk.edu.pl

Kostiantyn Yefremov

World Data Center for Geoinformatics and Sustainable Development, Kyiv, Ukraine  
k.yefremov@wdc.org.ua

Sergii Telenyk

Department of Theoretical Electrical Engineering and Computer Science, Faculty of Electrical and Computer Engineering, Cracow University of Technology Cracow, Poland  
stelenyk@pk.edu.pl

Volodymyr Khmeliuk

Department of Automation and Control in Technical Systems National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” Kyiv, Ukraine  
hmelyuk@gmail.com

**Abstract**—The approach to applications integration for World Data Center (WDC) interdisciplinary scientific investigations is developed in the article. The integration is based on mathematical logic and artificial intelligence. Key elements of the approach - a multilevel system architecture, formal logical system, implementation – are based on intelligent agents interaction. The formal logical system is proposed. The inference method and mechanism of solution tree recovery are elaborated. The implementation of application integration for interdisciplinary scientific research is based on a stack of modern protocols, enabling communication of business processes over the transport layer of the OSI model. Application integration is also based on coordinated models of business processes, for which an integrated set of business applications are designed and realized.

**Index Terms**—research, application integration, business processes, mathematical logic, formal logic, inference mechanisms, multi-agent systems, protocols, software agents

## I. INTRODUCTION: PARTICULARITIES OF WDC APPLICATION INTEGRATION

IN VIEW of the globalization of the economy and social life, National Science must integrate into the world and European organizations that promote the consolidation of research and consequently the development of scientific activity [2]. This is a very important process since there is an urgent need for interdisciplinary research, primarily for the assurance of sustainable development globally and regionally [3]. However, effective implementation of interdisciplinary research requires the creation of appropriate conditions for information exchange in the process of solving scientific problems. The scientific and technical progress that in its time facilitated the creation of information and communication technologies (ICT) nowadays benefits greatly from them. They are developing rapidly, covering new spheres of human activity and enhancing performance. Yet only field specialists can use ICT in a rational way, whereas the need for efficient information exchange within interdisciplinary research can only be met through rational ICT [19] application by specialists with deep knowledge in their areas of expertise [4].

The development of the information technologies (IT) domain that is experiencing qualitative changes related to ICT

Presented results of the research, which was carried out under the theme No. E-3/586/2018/DS, were funded by the subsidies on science granted by Polish Ministry of Science and Higher Education.

strengthening has created conditions for distributed computation and the efficient use of information and other resources. Consolidation of resources and the introduction of virtualization technologies are contributing to the process of substituting local solutions with distributed ones that allow the comprehensive use of all computing powers and data storage systems linked into a global network, thus granting access to accumulated information resources. The service approach formed on the basis of communication services has spread to the infrastructure, software development tools, and applications. The emergence of a wide range of new types of services, especially content-based ones, has led to the convergence of services and the formation of a generalized concept of information and communication services (ICS). The number of ICS providers has grown rapidly and convergent providers have emerged. The wide functionality, high quality, and moderate price of new services rendered by providers allow businesses to abandon the development of in-house or IT infrastructure and to use a wide variety of available ICS for component-based design of their information and telecommunication systems (ITS).

However, unified access to services is becoming a condition for the efficient use of the advantages of distributed systems and the possibilities of service-oriented technologies. At the same time, the formation of a new democratic IT environment, in which even small businesses can render services, has naturally been accompanied by the use of various tools and access technologies [5]. Therefore, historically the IT environment is heterogeneous, hence user access to its resources is to some extent complicated or at least inconvenient.

The same situation is characteristic of scientific activity in particular. For example, the World Data Centers (WDC) system created in 1956 under the aegis of the International Council for Science (ICSU) ensures collection, storage, circulation, and analysis of data obtained in various science areas [2]. During its existence, the WDC system has accumulated a lot of data and applications that may be used to solve the challenging problems of social development. They are one of the most powerful information resources used by hundreds of thousands of scientists, and the demand for it is increasing in proportion to the need for interdisciplinary research related to sustainable development, the solution of

urgent environmental protection issues, etc. However, problems related to the incompatibility of legacy applications caused by architecture differences, the variety of data presentation formats, and other factors prevent the effective use of such WDC resources.

Promising architectural solutions are being developed and gradually implemented in the ICT field, such as Next Generation Network (NGN) and Next Generation Service Overlay Network (NGSON) [6], that ensure the interaction of various transport layer technologies. Yet there is a need for a comprehensive integration of resources from various sources, and ICT developers' efforts should seek to enable scientists working in various domains to use the accumulated resources based on their areas of expertise, and not on the IT particularities. The sources, comprehensive access to which it is reasonable to ensure, include databases, websites and portals, various legacy file management systems, and data repositories structured according to various models.

Nowadays there exist more than 50 WDCs that for more than 50 years have created a system for data accumulation, analysis, processing, and international exchange. WDCs' powerful data storage systems retain huge volumes of astronomical, geophysical and other scientific data. WDCs' servers process this data using numerous and various applications powered by different technologies.

Certainly, from the point of view of scientists working on various resource-intensive problems, it is important to have access not to a large disordered system of possibilities, but to an integral complex of data and application sources intended to meet their specific needs, with a user-friendly interface that does not require any special IT knowledge. Nevertheless, the system for WDC data accumulation, analysis, processing, and international exchange does not provide such access. Furthermore, it was not designed for the growing level of scientific society's requirements and is not versatile enough to be used in interdisciplinary research. Therefore, a new interdisciplinary structure was created in 2008 — the World Data System (WDS) — to develop and implement a new coordinated global approach to scientific data, which guarantees omni-purpose equal access to quality data for research, education, and decision making. The new structure will have to solve the accumulated tasks, primarily the unification of formats and data transfer protocols, assurance of convenient access to data, and the organization of scientific data quality control [2].

Integrating formerly independent systems for the accumulation, storage, and processing of WDC data on a new advanced integration basis will allow a considerable enhancement of their overall efficiency. The creation of such a system will provide scientists with convenient centralized access to formerly separate resources, facilitating and quickening scientific and research activity around the world.

## II. RELATED WORK

An overview of existing integration solutions in all the mentioned aspects, starting with the integration technology aspect has been done. The existing solutions are powered by technologies for the creation, functioning, and development of distributed systems. Notwithstanding the differences of various technologies for the creation of distributed service-oriented systems, the overall principles and theoretical and methodological approaches are always similar. Independent services should be registered, described, and provided with the possibility to communicate transparently with clients and

with each other. Furthermore, networking interaction requires us to determine protocols for all levels of the OSI model. To do so, the corporate ITS standard structure with services intellectualization operations may be used, as suggested in [8], taking into account international, state, and branch standards, and corporate documents, first of all [9]. The structure consists of two parts, where the first covers the traditional four levels of standards of the TCP/IP protocols stack, and the second covers the user applications in accordance with the ITS class, destination, and services intellectualization operations. Figure 1 presents an example of the corporate standard structure.

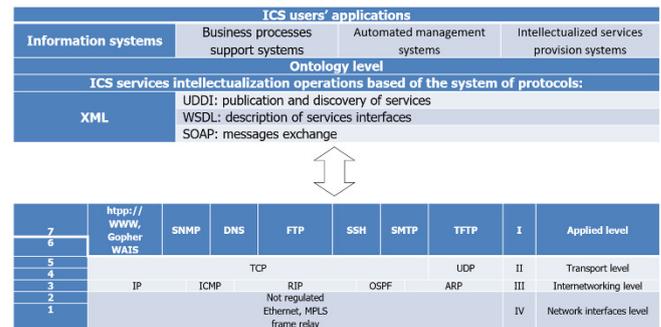


Fig. 1 Stack of services interaction protocols

For convenience, the figure presents the levels of the international standard for interaction of open OSI systems and their correlation with the corresponding levels of the TCP/IP protocols stack. The ITS classification used is proposed in [8]. It allows the systematization of various ICS by the level of users' requirements and queries, practical needs, and professional training, with no limitations imposed on the ITSs' functionality, attributes, or operations. The changes introduced apply to the first part of the structure and to the detailing of protocols that power the intellectualized interaction of applications in the course of solving users' more complex problems.

The services registry is maintained by the Universal Description, Discovery and Integration (UDDI) technology [10] that allows both people and client-programs to publish information on services and search for a required service.

The Web Service Description Language (WSDL) that, according to the W3C definition, constitutes an XML format for the description of networking services as a set of operations working with document- or procedure-oriented information through messages [11], is used for the unified description of services, which allows them to be used independently from the programming language. WSDL documents, by virtually creating a unified layer that allows the use of services created on the basis of various platforms, describes the service interface, URL, communication mechanisms 'understood' by the service, methods provided by it with corresponding parameters (type, name, location of the service Listener), and the service messages structure.

To implement the key part of the interaction—messages exchange—one of the widespread technologies may be used, and the selection should be determined by compliance with the requirements of the system for data accumulation, processing, and exchange:

**SOAP:** a user-friendly technology that is easy to use with the Business Process Execution Language (BPEL), which ensures interaction of distributed systems irrespective of the object model, operating system, or programming language. Data is transferred as special-format XML documents [12];

**CORBA:** a mechanism created to support the development and deployment of complex object-oriented applied systems, which is used for the integration of isolated systems, allowing programs that are developed in different programming languages and working in different network nodes to interact with the same ease as if they were in the address space of a single process. Such interaction is ensured through unified construction of their interfaces using a special-purpose declarative Interface Definition Language (IDL). At the same time, the interface and the description thereof do not depend on the operating systems or the processor architecture either [13];

**REST** (Representational State Transfer) is a style of software architecture for distributed systems, which is used to create web services. A global URL identifier unambiguously identifies each unit of information in an unvarying format. Data is transferred without any layers [14];

**WCF** (Windows Communication Foundation) is Microsoft's platform designed to create applications that exchange data through the network and are independent from style and protocol [15].

Although nowadays the most acceptable choice seems to be SOAP, based on which the interaction presented in Fig. 1 is performed, all the technologies allow the integration of services of various origin [16] – [17].

To take into consideration service-based, process-functional, and component-based approaches to the design and maintenance of users' solutions, the following well-known architectural means of organizing shared functionality are used:

Web Service Choreography is the approach that determines the protocols of web services' interaction to perform a single global task by performing parts thereof. The role assumed by the service determines its model of exchanging messages with other services. This method shows high efficiency for small tasks, but with increased complexity of tasks the number of services involved grows rapidly, solutions become too massive, and the efficiency drops quickly [18].

Based on the abovementioned technologies, a few solutions have been worked out, which may be used to create applied systems for the accumulation, processing, and exchange of scientific data in different areas. The best known among them are the ESIMO and GEOSS systems. They are quite widespread, although they have a number of drawbacks in terms of processing heterogeneous information.

One of the most crucial issues is that of developing mathematical models, methods, and means for the integration of various applications, both legacy ones based on traditional technologies and client-server and web-oriented technologies. The article proposes an approach to applications integration using the mathematical logic instrument and artificial intelligence theory for interdisciplinary research through the example of the WDC system functioning. The specificity of the applications integration for interdisciplinary research is

that the coordination of business processes models is not required because it is, in fact, substituted with schemes for performing users' tasks. In general, the applications integration is performed on the basis of coordinated business processes models, and the integrated complex of business applications is intended to support them.

### III. DEFINITION OF THE APPLICATION INTEGRATION PROBLEM FOR INTERDISCIPLINAR REASARCH

Mathematical models and methods as the basis for a holistic solution should be created to power data storage and processing centres, which will provide users with versatile possibilities at the level of information integration and servers availability. The mentioned models and methods should be devoid of the drawbacks characteristic of the algorithmic approach, related to the need to reprogram data processing algorithms upon the emergence of new data types and changes in the implemented algorithms or the emergence of new ones. At the data processing level, such a solution should ensure:

- computations distribution and the use of remote hardware resources;
- versatility and adaptation to the system load;
- system usage space;
- data supply by remote client or service;
- availability of intellectual data processing means directly to the end user with no special knowledge or skills;
- fast and simple integration of data and applications of various global information systems

### IV. THE APPLICATION INTEGRATION SYSTEM ARCHITECTURE

The principal tenet for the creation of a distributed system is the method of organizing services interaction. Of the two known main ways of services interaction - orchestration and choreography - the more efficient for the WDC systems is the former. Indeed, orchestration that aggregates basic services into hierarchically integrated systems, subordinating them to administrators – 'orchestrator' services - allows services to be unified by attributes that are convenient to WDC (science area, functionality, regional location, etc.), and to provide orchestrator services with powers in accordance with the international data exchange policies. Thus, it is possible to line up a simple and effective system, in which the search for the necessary services or the construction of their composition for the complex queries inherent in interdisciplinary research will not require a lot of time, as every Orchestrator Service has information on the functionality of inferior services, and in addition they can coordinate their possibilities in the process of the planning and execution of users' queries. Plugging new services into the system will not present a particular problem either, as doing so will only require the service description to be laid down, entered into the registry, and assigned to a certain orchestrator service.

Such an approach will work when users know exactly their information needs and have corresponding knowledge and skills to compile chains of queries to the known orchestrators. What is more important is that the approach will provide the opportunity to work with the system for users who cannot initiate services, by determining the sequence of their work and specifying execution and interaction parameters. Users only have to know how to formulate their needs in terms of a particular subject domain. In this case, the association of services and the organization of their cooperation require an intellectual constituent. For this purpose, intellectual agents are used, which constitute the system core, implementing its functioning logic, which promotes the formation of queries, plans their execution, and organizes basic services interaction. The introduction of intellectual agents as orchestrator services into the hierarchical structure forms a two-tier system; the bottom level consists of services performing basic tasks, the top level consists of intellectual agents that orchestrate basic services. By using the registries of subordinated basic services and their functionality and by interacting with each other, interconnected intellectual agents implement methods of logical inference. The inference result is the composition of basic agents' operations that allow user-defined tasks to be performed. This operations composition, or proof, is transferred to the lower level, where the operations necessary to solve the user's task are performed. At this point, control is handed over to the lower level agents that only return the final result to be sent to the user.

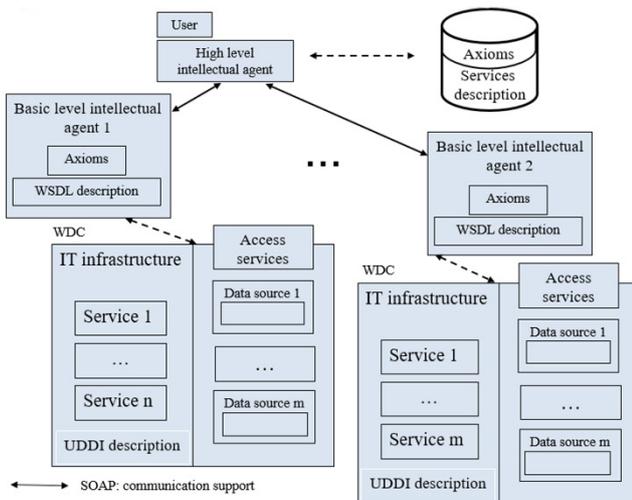


Fig. 2 System components interaction scheme

The user's query itself is performed upon formation of the proof containing references to one or several pointers to data sources and IDs of methods that the system services should apply to the data. Some queries do not require pre-processing of data, since the execution of services methods is sufficient upon receipt of information from the data source. Others require a certain sequence of prior data operations, and as a result, the execution of other services methods. The necessary connections are described in the axioms entered into the knowledge base upon registration of respective services in the system. These operations yield the final product of the system - data that are the solution to the user's problem. Since the system operation is determined by the user's query, the

system interface should help them make a query in familiar terms.

The proposed solution uses UDDI to create the service registries and WSDL for the unified description thereof. The key part—messaging—is implemented using SOAP, and services orchestration—using BPEL. Agents are described by means of JADE. The system components interaction scheme is presented in Fig. 2.

The description of the interface (WSDL) of every low-level service registered in the system and of related axioms is input into the system knowledge base and directly into the knowledge base of the Controller Agent to which the new service will be subordinated. The Controller Agent is selected in accordance with the policies adopted in the WDC system, for example, by the criteria of the geographical location of the deployed agent, in order to minimize data exchange.

At the agent level, the orchestrator agent (or several), having received the user's query, interrogates the agents in order to search for the necessary services and available resources. The user's query is processed taking into account the services of the system, their functionality (described by the axioms of the system), and the inference rules defined by logical formalism. The resulting proof is transferred to the lower level to be implemented.

At the services level, each agent invokes the required services from its set and sends processed data according to the action chains. This process is carried out in accordance with the task solution tree reconstructed by the solution tree reconstruction mechanism based on the proof.

To realize this interaction of services in the WDC system, the most appropriate is the logical approach, which allows both the level of the abovementioned requirements to be reached and the drawbacks of a traditional algorithmic approach to be eliminated. Indeed, it is only required to develop the inference method and the solution tree reconstruction mechanism that will implement the query formation processes, plan their execution, reconstruct and implement the task solution scheme. The logical approach is the most appropriate one to create and describe these constituents. The logical approach implementation requires one to:

- describe the existing applications and their functional capacities in the formal language;
- formulate the inference rules;
- determine the inference method;
- develop an algorithm of the user's query execution tree reconstruction based on the proof;
- implement these methods and mechanisms in the agents of the system.

## V. THE FORMAL LOGICAL SYSTEM

Let us describe the formalism upon which the program system that will ensure solution of the formulated problem will be built. We will take the first order clausal logic for a base and describe the formal system language in accordance with the structural elements determined in [3].

**Symbols: service:** ( ), [ ], { }, :, <, >, ;

**constant:**

- 1) *individual, of primary types* (int, real, char, bool) -  $a_1^1, a_2^1, \dots, a_1^2, a_2^2, \dots$  where each constant  $a_i^k$  pertains to type (primary type)  $k$ ; *structural type* (construct) -  $c_1, c_2, \dots$ ; *procedural type* (method) -  $d_1, d_2, \dots$ ; *objective type* (problem, entity, relation) —  $e_1, e_2, \dots$ ;
- 2) *functional i-place*, for individuals of type  $k$  -  $h_1^1, h_2^1, \dots, h_1^2, h_2^2, \dots$ ;
- 3) *predicate i-place*, for individuals of type  $k$  -  $A_1^1, A_2^1, \dots, A_1^2, A_2^2, \dots$  (this class includes taxonomic, relational and other predicates, as well as traditional relations, at least equality = and order  $\geq$ );

**variable:** for individuals of type  $k$  -  $x_1^1, x_2^1, \dots, x_1^2, x_2^2, \dots$ ,

where every variable  $x_i^k$  pertains to type  $k$ ;

**logical:**  $\neg, \wedge, \vee, \leftarrow, \exists, \forall, \leftrightarrow$

**Individual terms of type  $k$ :**

- 1) each individual constant  $a_i^k$  of type  $k$  is an individual term of type  $k$ ;
- 2) each free variable  $x_i^k$  for individuals of type  $k$  is an individual term of type  $k$ ;
- 3) if  $h_i^j$  is a certain functional constant for individuals of type  $k$  and  $\tau_1, \dots, \tau_j$  are terms for individuals of type  $k$ , then  $h_i^j(\tau_1, \dots, \tau_j)$  is an individual term of type  $k$ ;
- 4) there are no other individual terms of type  $k$ .

The terms obtained by applying construction rules 1 or 2 of the definition will be called primary, and all the others—complex.

**Formulas for individuals:**

- 1) if  $A_i^j$  is a predicate constant for individuals and  $\tau_1, \dots, \tau_j$  are terms for them, then  $A_i^j(\tau_1, \dots, \tau_j)$  is the atomic formula for individuals;
- 2) the atomic formula for individuals is the formula for them;
- 3) there are no other formulas for individuals.

Hereinafter we consider that the system contains an omni-purpose transformer of formulas into the traditional for the clausal form (we are talking about the Horn clauses) view with a single  $\rightarrow$  symbol, atomic formulas to its left and right, and an implicit quantifier  $\forall$ .

Specifiers, preconditions, post-conditions, specifiers of methods, specifiers of problems, clause, the system's knowledge and inference rules are presented in detail in [1].

## VI. THE INFERENCE METHOD

The method proposed in [2] based on analogy and types of assertions has been used. A detailed description of this method can be found in [1], [4], [7].

The inference mechanism work algorithm is presented in Fig. 3.

To improve the inference mechanism three known elements were integrated:

- 1) a multiset of literals (atomic formulas) (which will be called a multi clause (m-clause));
- 2) the ordered linear;
- 3) typification abstraction (to manage an ordered linear proof).

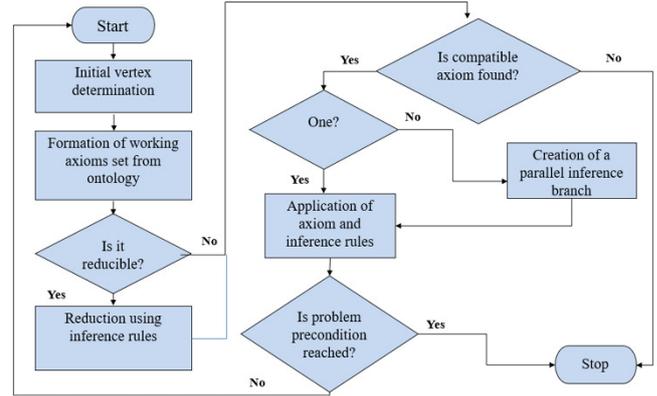


Fig. 3 Inference mechanism work algorithm

## VII. THE MECHANISM OF SOLUTION TREE RECONSTRUCTION

To obtain from the proof generated by the inference mechanism a functional sequence of actions to be used by our system, taking account of services features and nature, the solution scheme reconstruction mechanism should be initiated [1].

Condition: proof Result = $\langle V, T \rangle$ . To find: $G = \langle V, E, \Theta \rangle$	
Step	
1	Identification of the terminal triple—the triple that contains vertex $k$ that is neither the first nor the second component of any of the triples of set $T$ .
2	In the selected triple, the initial vertex is the post-condition vertex, the second is the method vertex, and the terminal is the precondition vertex. The method that turns the precondition into the post-condition is determined by the unambiguous correspondence to the vertex formulas of the post-condition and precondition of the method axiom. The vertices are connected with the edges: precondition to method and method to post-condition. If the precondition of the method axiom consists of a combination or intersection of several elements, a data vertex is inserted between the precondition and the method vertices, which groups the preconditions according to the relevant inference rule. If the terminal triple only contains two vertices, it means that the inference rule was used to convert the vertex clauses, and the place of the method vertex in this tree branch is occupied by the transformation data vertex that corresponds to the inference rule used, with the corresponding post-condition vertices.
3	The vertices, edges, and their correspondence are introduced into $G$ .
4	Having extracted the processed triple from the set of triple vertices $T$ , we repeat step 2. If the set is empty, the tree reconstruction is completed.

Fig. 4 Algorithm for vertex scheme reconstruction

The solution scheme constitutes a connected directed graph with no oriented cycles with parallel directed paths from the root to the vertices; it has three types of vertices, and is specified by triple  $G = \langle V, E, \Theta \rangle$ , where  $V = V_1 \cup V_2 \cup V_3$ ,  $V_1$  is a set of method vertices,  $V_2$  is a

set of precondition vertices and post-condition vertices,  $V_3$  is a set of data vertices in which data is merged or split;  $E$  is a set of edges;  $\Theta$  is a subset of Cartesian product  $E \times V \times V$ , which determines the correspondence of edges to pairs of vertices. The scheme determines the sequence and the correspondence according to the data of actions to be performed by the data processing system's executive mechanisms to obtain the result desired by the user [1].

Algorithm for vertex scheme reconstruction was presented on Fig. 4.

The reconstructed solution tree is fed to the actuator input. The implementation of a particular method specified in the solution tree is represented by a construct that describes the input data (entities, connections, relations between them), method preconditions, method post-conditions, and output data. The solution tree branches downstream of the data-splitting vertex can be executed in parallel until they reach the data-merging vertex, where, after completing all the parallel branches involved in merging, they continue to be executed consecutively. The work algorithm of the solution tree reconstruction mechanism is presented in Fig. 5.

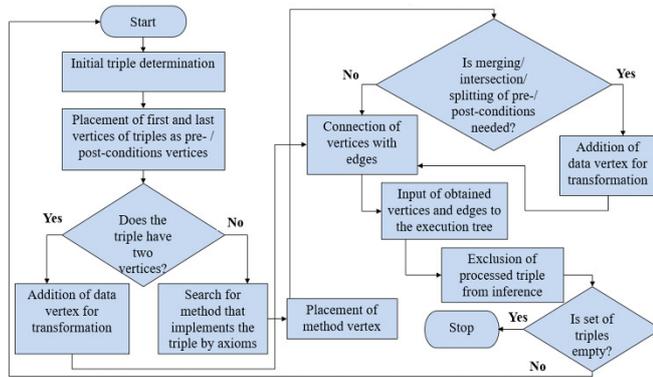


Fig. 5 Work algorithm of solution tree reconstruction mechanism

## VIII. APPLICATION OF LOGICAL APPROACH TO PROBLEM SOLUTION

The approach efficiency on the real scientific task of calculating the component of life safety and indicating critical values of threat indicators for analyzing the sustainable development of the regions of Ukraine has been demonstrated.

The life safety component formula:  $Csl = \sqrt[3]{\sum_0^n Threat^3}$ , where the threat value is normalized by formulas (2) or (3). The purpose of indicating critical values of threat indicators is to determine the priority in consideration thereof in the decision-making process on the level of a single region and the entire country in order to mitigate the impact of threats on sustainable development.

Suppose that for every administrative unit  $i = \overline{1, n}$  there is a set of values  $\langle x_{i,1}, x_{i,2}, \dots, x_{i,m} \rangle$  of indicators  $X_j, j = \overline{1, m}$ , which characterize the negative impact of certain phenomena on the sustainable development processes in the economic, social, and ecological spheres. Such indicators, whose essence and

composition are determined by experts, will be called threat indicators.

Given the content of the critical values indication problem, the following characteristic function must be determined:

$$\Psi(x_{i,j}) = \begin{cases} 0, & \text{if value of } x \text{ is not critical} \\ 1, & \text{if otherwise} \end{cases} \quad (1)$$

where  $i = \overline{1, n}, j = \overline{1, m}$ .

It is clear that the determination of function  $\Psi(x_{i,j})$  must be based upon certain criteria that take into consideration exceeding by  $x_{i,j}$  a hazardous limit, the relative position of region  $i$  in the indicators rating  $X_j$  compiled for the comparison groups and for the entire country, and the degree of "hazard" of value  $x_{i,j}$  in comparison to values of other indicators for region  $i$ .

To account for the relative position of the region in the entire country's ratings, the following criterion is used:

$$R_{i,j} = \left( 1 + e^{\frac{a - x_{i,j}}{b}} \right)^{-1} \quad (2)$$

if higher values of indicator  $X_j$  correspond to a higher impact of the respective threat on the sustainable development, and:

$$R_{i,j} = 1 - \left( 1 + e^{\frac{a - x_{i,j}}{b}} \right)^{-1} \quad (3)$$

if lower values of indicator  $X_j$  correspond to a higher impact. In formulas (2)-(3), parameters  $a$  and  $b$  are calculated by the following formulas:

$$a = \overline{X_j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, \quad b = \sigma(X_j) = \sqrt{\frac{\sum_{i=1}^n (x_{i,j} - \overline{X_j})^2}{n}} \quad (4)$$

Criterion  $R_{i,j}$  is a dimensionless number that assumes values within  $[0, 1]$ . Values of around 0.5 correspond to average values of  $X_j$  in the selection, and values higher than 0.75 correspond to values that exceed the average ones by more than a standard deviation. In this case, the characteristic function (1) taking account of one criterion  $R_{i,j}$  may be expressed as follows:

$$\Psi_R(x_{i,j}) = \begin{cases} 0, & R_{i,j} < 0,75; \\ 1, & R_{i,j} \geq 0,75. \end{cases}$$

Criterion  $P_{i,j}$  that takes into account a region's relative position in the comparison group may be calculated by formulas (2)-(3) taking account of the fact that parameters  $a$

and  $b$  are calculated by formula (4) independently for each comparison group.

Criteria  $R_{i,j}$  and  $P_{i,j}$  are dimensionless numbers and are of similar nature and can therefore be aggregated through a weighted sum:

$K_{i,j} = w_R R_{i,j} + w_P P_{i,j}$ ;  $w_R + w_P = 1$ , where weighting factors  $w_R$  and  $w_P$  are determined by experts.

Thus, for region  $i = \overline{1, n}$  we have a set  $\langle K_{i,1}, K_{i,2}, \dots, K_{i,m} \rangle$  of values of the aggregated criterion that accounts for a region's relative position in ratings compiled for comparison groups and for the entire country. Now, among values  $K_{i,j}$ ,  $j = \overline{1, m}$ , the worst must be determined, which may also be performed by formula (2):

$$I_{i,j} = \left( 1 + e^{\frac{a - K_{i,j}}{b}} \right)^{-1}$$

where parameters  $a$  and  $b$  are calculated in selection  $K_{i,j}$ ,  $j = \overline{1, m}$ .

For values of criterion  $I_{i,j}$  the same remarks apply as for criterion  $R_{i,j}$ . Therefore, characteristic function (1) may be expressed as follows:

$$\Psi_I(x_{i,j}) = \begin{cases} 0, & I_{i,j} < 0,75; \\ 1, & I_{i,j} \geq 0,75. \end{cases}$$

Thus, values  $\Psi_I(x_{i,j}) = 1$  correspond to the highest priority of attention to be paid to the value of indicator  $X_j$  in the administrative decision-making process on the level of single region  $i$ .

## IX. CONCLUSIONS

Based on the analysis of existing data centres, their equipment and software, a high quality solution is offered that provides a simple and flexible way to integrate heterogeneous information systems and their services into the World Data System. One of the key features of the proposed solution is the automation of the algorithm construction of the actions sequence that executes users' queries. A logical formalism has been created to describe this solution, and on its basis an inference method and a solution tree reconstruction mechanism have been developed.

The analysis of available technologies used for the implementation of distributed systems allowed the use of such a set of software solutions for practical implementation of this solution: UDDI for creating a registry of services entered into the system; WSDL for a unified description of services; SOAP for exchanging notifications between services; BPEL for the overall coordination of services. Intellectual agents can be implemented using JADE.

The implementation of the proposed solution will provide an opportunity to use all the integrated computing capacities and data storage systems of the World Data System in a comprehensive manner. Thus, users will be able to easily gain access to all the necessary resources and services available to the system.

## REFERENCES

- [1] S. Telenyk, G. Nowakowski, K. Yefremov and V. Khmeliuk, "Logics based application integration for interdisciplinary scientific investigations", 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 1026-1031, Bucharest, 2017. DOI: 10.1109/IDAACS.2017.8095241
- [2] M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny, "Integration of the Ukrainian science into the world data system", Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010), pp. 211-219. DOI: 10.1007/s10559-010-9199-9
- [3] M. Z. Zgurovsky, A. O. Boldak, K. V. Yefremov and others, "Analysis of Sustainable Development – Global and Regional Contexts", International Council for Science (ICSU) and others. – K.: NTUU «KPI». – Part 2. Ukraine in Sustainable Development Indicators (2011-2012). – 232 p, 2012.
- [4] O. Pavlov, S. Telenyk. *Algorithmization and IT in management*, Kyiv: Technics, 2002 – 320 p.
- [5] Data Integration Information, quick view on world of data, (online) homepage at: <https://www.dataintegration.info/>
- [6] M. Ulema et al., "Next generation service overlay networks (NGSON)", IEEE Communications Magazine 50(1):52-53, 2012, DOI: 10.1109/MCOM.2012.6122532.
- [7] A.Y. Levy, "Logic-Based Techniques in Data Integration", In: Logic Based Artificial Intelligence, Edited by J. Minker. Kluwer Publishers, 2000.
- [8] P. P. Maslianko, "Fundamentals of the methodology of system design of information and communication systems", Naukovi Visti NTUU "KPI," No. 6, 54-60 (2007).
- [9] OMG Systems Modeling Language, (online) homepage at: <https://www.omg.org/spec/SysML/About-SysML/>
- [10] UDDI Version 3.0.2 Specification, (online) homepage at: [http://uddi.org/pubs/uddi\\_v3.htm](http://uddi.org/pubs/uddi_v3.htm)
- [11] J. Greer, "Web Services Description Language: 55 Most Asked Questions: What You Need to Know", Emereo Publishing, 2014
- [12] SOAP Version 1.2 Part 1: Messaging Framework (Second Edition), (online) homepage at: <https://www.w3.org/TR/soap12-part1/>
- [13] About the common object request broker architecture specification version 3.3, (online) homepage: <https://www.omg.org/spec/CORBA>
- [14] G. Nowakowski, "Rest Api safety assurance by means of HMAC mechanism", Information Systems in Management, Vol. 5, No. 3, pp. 358-369, 2016.
- [15] Windows Communication Foundation Architecture Overview, (online) homepage at: <http://msdn.microsoft.com/en-us/library/aa480210.aspx>
- [16] S. Graham, et al., "Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI (2nd Edition)", Sams Publishing; 2 edition, 2004
- [17] S. El-Scoud, H. El-Sofany, M. Abdelfattah, M. Reham, "Big Data and Cloud Computing: Trends and Challenges", International Journal of Interactive Mobile Technologies, 2017, Vol. 11 Issue 2, p34-52, 19p, 3 Diagrams; DOI: 10.3991/ijim.v11i2.6561
- [18] J. Laznik, Y. Mannari, R. Dhruv, BPEL and Java Cookbook: Over 100 Recipes to Help You Enhance Your SOA Composite Applications with Java and BPEL, Birmingham, 2013
- [19] E. Ziemba, "The ICT Adoption in Government Units in the Context of the Sustainable Information Society", 2018 Federated Conference on Computer Science and Information Systems, pp.725-733. DOI: 10.15439/2018F116