УДК 004.4 DOI: 10.31673/2412-4338.2022.049098

Ruzhinsky V.G., Bondarchuk A.P.
State University of Telecommunications, Kyiv

Kuzminykh V.O., Otrokh S.I., Taranenko R.A.
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv

EVENT-ORIENTED ARCHITECTURE SYSTEM FOR COLLECTING AND PROCESSING BIBLIOGRAPHIC INFORMATION

Abstract: *Today, one of the most important indicators of the level of international activity of a scientific institution is the number and quality of scientific publications placed in the most prestigious international bibliographic databases. The collection and processing of such information requires significant time and resources, given their size and daily replenishment, the variety of storage formats and access to information repositories. Therefore, such a task of collecting, processing, consolidating and analyzing bibliographic information and assessing the level of international activity of researchers of state scientific organizations and teachers of higher educational institutions is a task of processing big data and requires special approaches in the construction of specialized software systems for its solution.*

The article considers the construction of the structure of the software system and the main elements of the event-oriented microservice architecture of the system of collecting and analyzing information from bibliographic sources. The basis for the implementation was the microservice architecture with the distribution of services according to the functional purpose. The proposed structure of the system makes it possible to adaptively change the selection of information sources based on the evaluation of the effectiveness of obtaining relevant information from relevant sources, in accordance with the events that occur during the collection of information. The choice of directions for data collection and processing is based on the use of typical scenarios, which allows to significantly increase the efficiency of the analyst's work. The event-oriented microservice architecture of the system makes it possible to adapt the operation of the system to the loads on individual microservices by analyzing the created relevant events.

The developed system provides an opportunity to collect information about the international activities of individual people or institutions from sources that are heterogeneous in terms of storage form and composition. The composition of information sources in this system can be expanded without interfering with the work of other system components. Each individual processor of information from the sources can provide only partial data, which is supplemented at the expense of others. The system adapts to the user's request in such a way that data is taken only from those sources that meet the request's requirements and can satisfy them.

The developed approach is built on the basics of event-oriented microservice architecture and can be used in the development of various information and analytical systems.

Keywords: *microservices, adaptation, event-driven architecture, information*

Ружинский В.Г., Бондарчук А.П.
Державний університет телекомунікацій, Київ

Кузьміних В.О., Отрох С.І., Тараненко Р.А.
Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ

ПОДІЙНО-ОРІЄНТОВАНА АРХІТЕКТУРА СИСТЕМИ ЗБОРУ ТА ОБРОБКИ БІБЛІОГРАФІЧНОЇ ІНФОРМАЦІЇ.

Анотація: *Сьогодні одним із найважливіших показників рівня міжнародної діяльності наукової установи є кількість і якість наукових публікацій, розміщених у найпрестижніших міжнародних бібліографічних базах даних. Збір та обробка такої інформації потребує значного часу та ресурсів, враховуючи їх розмір та щоденне поповнення, різноманітність форматів зберігання та доступ до інформаційних сховищ. Тому таке завдання збору, обробки, консолідації та аналізу бібліографічної інформації та оцінки рівня міжнародної діяльності дослідників державних наукових організацій і викладачів вищих навчальних закладів є завданням обробки великих даних і потребує особливих підходів у побудові спеціалізованих програмні системи для її вирішення.*

У статті розглянуто побудову структури програмної системи та основні елементи подійно-орієнтованої мікросервісної архітектури системи збору та аналізу інформації з бібліографічних джерел. Основою для реалізації стала мікросервісна архітектура з розподілом сервісів за функціональним призначенням. Запропонована структура системи дає можливість адаптивно змінювати відбір джерел інформації на основі оцінки ефективності отримання релевантної інформації з релевантних джерел, відповідно до подій, що відбуваються під час збору інформації. Вибір напрямів збору та обробки даних базується на використанні типових сценаріїв, що дозволяє значно підвищити ефективність роботи аналітика. Подійно-орієнтована мікросервісна архітектура системи дозволяє адаптувати роботу системи до навантажень на окремі мікросервіси шляхом аналізу створюваних відповідних подій.

Розроблена система дає можливість збирати інформацію про міжнародну діяльність окремих осіб чи інституцій із джерел, неоднорідних за формою та складом зберігання. Склад джерел інформації в цій системі можна розширювати, не втручаючись у роботу інших компонентів системи. Кожен окремий обробник інформації з джерел може надати лише часткові дані, які доповнюються за рахунок інших. Система адаптується до запиту користувача таким чином, що дані беруться тільки з тих джерел, які відповідають вимогам запиту і можуть їх задовольнити.

Розроблений підхід побудований на основах подійно-орієнтованої мікросервісної архітектури та може бути використаний при розробці різноманітних інформаційно-аналітичних систем.

Ключові слова: *микросервіси, адаптація, подійно-орієнтована архітектура,інформація*

1. Introduction

The number of sources of scientific and technical information and their volume is constantly growing from year to year, and therefore an effective method of its search, collection and processing is a key factor in the scientific and technological development of both individual scientific teams and humanity in general. The number of sources of information available to mankind and their total volume have long exceeded the capabilities that humans have for processing and control. Therefore, without special information, both software and technical means, the opportunity to fully use and analyze such volumes of modern scientific and technical information is lost.

One of the fairly common and important areas of scientific information analysis is the analysis of the current activity of scientific teams and individual scientists based on the results of their publications, which are reflected in specific recognized scientific bibliographic databases. Such information is the basis for assessing the level of development of scientific and elementary institutions, individual scientific groups and individual scientists for their further comparison and makes it possible to forecast the efficiency indicators of the state of their scientific activity. According to the state of these indicators, scientific activity is evaluated both within the framework of a separate institution, the country, and at the international level in comparison with similar institutions.

Such assessment takes place in the form of formation and analysis of various rating systems based on defined sets of indicators of scientific activity. This approach provides an opportunity to assess the current situation and predict the value of important influential indicators of the effectiveness of scientific activity for the future. In addition, it becomes an important factor in determining the scientific position of the institution in the country and in the world, which is important for the formation of promising directions of development and financing of international activities of institutions, groups of scientists and individual scientists.

The international activity of universities and research institutes, as a permanent concept, does not currently have a clear definition. In most cases, depending on the goals and objectives set for themselves by the interested party or the customer in such an analysis, both the number of criteria and their indicators used for evaluation, and the method of calculating the complex criterion will depend.

There are quite a large number of indicators of the level of international activity, such as the number of foreign students and teachers, international educational programs, internships of specialists abroad, etc. But today, one of the most important indicators of the level of a scientific organization, a group of scientists, or a specific individual scientist is the number of scientific publications, the level of domestic and foreign publications, where they are published, references to them in priority international bibliographic databases, and the number of citations of scientific publications placed in international bibliographic databases. At the same time, as a rule, the most important indicator is the level of international bibliographic bases.

Information that accumulates knowledge about scientific publications can be found in a large list of global bibliographic sources, which are regularly replenished with new scientific works every day. Information about other types of scientific activities of institutions, especially international, is difficult to access, is not consolidated in a specific resource and is not, in most cases, sufficiently complete and up-to-date. Therefore, consolidated information that describes the international activity of scientists based on the analysis of their publications can give a more or less accurate picture for assessing the comparative level of their international activity. The task of information consolidation is one of the most important tasks of processing large volumes of data [1].

To solve this problem, it is necessary to build a software system that, based on the analysis of large data obtained from open bibliographic systems and auxiliary local databases for storing bibliographic information, collect, process, consolidate and analyze information in order to provide conclusions about the comparative levels of international activity as scientific organizations as well as individual scientists.

The main tasks that the system will solve are:

– collection of data from sources different in structure and composition, mainly bibliographic and abstract scientometric databases,

– bringing to a single form and aggregating the received data in accordance with the user's request,

– removal of redundant erroneous data and duplicates of already received data.

A system aimed at solving this kind of problems should take into account the possibility of using a variable number of information sources, which will differ among themselves both in terms of data structure and interaction interfaces. At the same time, the main requirement is the creation of a flexible architecture with the possibility of its constant modification by including new sources to the system and processing functions, modification of existing functions as needed.

The sources used to collect such information, from the point of view of the task of collecting and processing information to assess the level of activity of scientific and elementary institutions, individual scientific groups and individual scientists for their further comparison, have two main features that affect the construction of the architect of software systems for collection and information processing. First, the number of items relevant to the query in the bibliographic database may be very small. Secondly, such information, or more precisely, each request to the source of information may have a certain cost.

Therefore, it becomes a task to build such a system that will make it possible to minimize the number of requests due to the use of information sources that have the largest number of records relevant to the request. As a rule, it is almost impossible to determine which of the sources used have the largest number of records relevant to the request, and therefore the adaptation of the software system to the request can be built on the basis of using an event-oriented approach to building the microservice architecture of the software system.

2. Choice of system architecture

Micro-service architecture is a further development and a separate type of service-oriented architecture and is built on separate isolated components - micro-services. The micro-service architecture of a software system consists of such components as services, service bus, external configuration, API gateways and containers. Properties of micro-service architecture include: separation into service components, needs orientation, product orientation, decentralized data management, infrastructure automation, protection against failures, intelligent endpoints. The breakdown of microservices is based on the approach of subject-oriented design DDD (Domain-driven design). Containers are created on the basis of templates, which represent an image formed by different levels.

The microservice architecture of building information systems is a complex of independently deployed services. The system built in this way has the ability to evolve in parts, primarily due to the fact that each microservice is largely autonomous. In addition, such an architecture allows flexible scaling of information system components, thereby ensuring optimal use of available server equipment. At the same time, each microservice is scaled independently of other microservices. It became possible to allocate computing resources taking into account real user activity and demand for one or another functionality of the system.

Microservice architecture [2,3] is an approach to creating an application that implies the rejection of a single, monolithic structure. This is one of the ways to develop software applications by creating separate software modules independent of each other. Each of them is responsible for a certain task, can be changed, supplemented and expanded.

A key feature of the system is the adaptation of system parameters according to the user's request. That is, depending on the request, one of the possible ways of extracting and processing data is formed, according to the features of the request and the capabilities of the system to process this request. This is achieved thanks to a generalized flexible data model, event-oriented, microservice architecture and orchestration of software services.

Event-oriented architecture (EDA) is a direction of software architecture that takes into account events that occur during the solution of a problem using microservices and reacts to them by influencing the management of microservices. An event is interpreted as some action that triggers corresponding messages or a change in the application. Thus, an event is a significant change in the state of the software system. In practice, event-oriented architecture is also considered as one of the variants of adaptive micro-service architecture. In programs of event-oriented architecture, the core of the system is the database. In EDA, the focus shifts to events and how they are handled by the system architecture. The logic of the event architecture is built according to two types of topology - Broker (Broker) and Mediator (Mediator), named by programs as mediators that connect the generator and consumer of events. The main elements of the topology: Event queues, Event mediator, Event channels, Event processors.

The main advantages of EDA are the output of the result in real time. lower data transmission delays, higher bandwidth, simple scalability; high fault tolerance and availability.

3. Implementation of system architecture

The basis for implementing the system was an adaptive event-oriented microservice architecture with the definition of services according to their functional purpose. This was necessary in order to be able to dynamically expand the functionality of the system without significant intervention in already existing elements of the system. This allows for greater flexibility in scaling to increase system efficiency and throughput.

This approach gives the following important [4]:

– resistance to failures when one of the services fails;
– the system can include applications in different programming languages;
– simple scalability;
– independence from changes of individual services;
– data security and control;
– the possibility of partial implementation to be provided to the customer;
– the possibility of developing parts of the system by separate distributed teams of developers.

The general functional structure of the elements of the software system (Fig. 1) includes a number of basic elements that reflect the main functions and connections between individual elements of the structure. The system is built on the basis of WEB access to the main work functions of users.

The management of the system interface distinguishes two main functional elements of the system:

– administration interfaces,
– analyst interfaces.

Administration interfaces perform the following main functions:

– granting access rights,
– system security policy implementation,
– connection of new software modules of the system,
– testing new software modules of the system,
– connecting new information sources,
– formation and connection of new search and data processing scenarios,
– testing and adding new scripts to the standard script repository.

Scenario management [5] within the framework of the system implements the functions of choosing a typical scenario or forming a new scenario both directly on the basis of access to information sources and on the basis of the system's functional capabilities for processing information and providing results in a specified form.

The results obtained from the sources of information specified in the scenario are pre-processed, checked for relevance, relevance, uniqueness and consolidated in the database of search results.

This information, after appropriate processing, is transferred to the analyst's workplace and displayed in the form specified in the data collection and processing scenario.

Further detailed processing of the received data, which is necessary for the analyst to further analyze the results, is carried out within the functions defined by the capabilities of the analyst's workplace.

Scopus, Core, arXiv, DBPL, Web are used as sources of information of Science and regional.

In addition, the data of the analysis includes data from internal information bases of scientific data of organizations, data from questionnaires on the activities of scientists, and data on participation in scientific contracts.

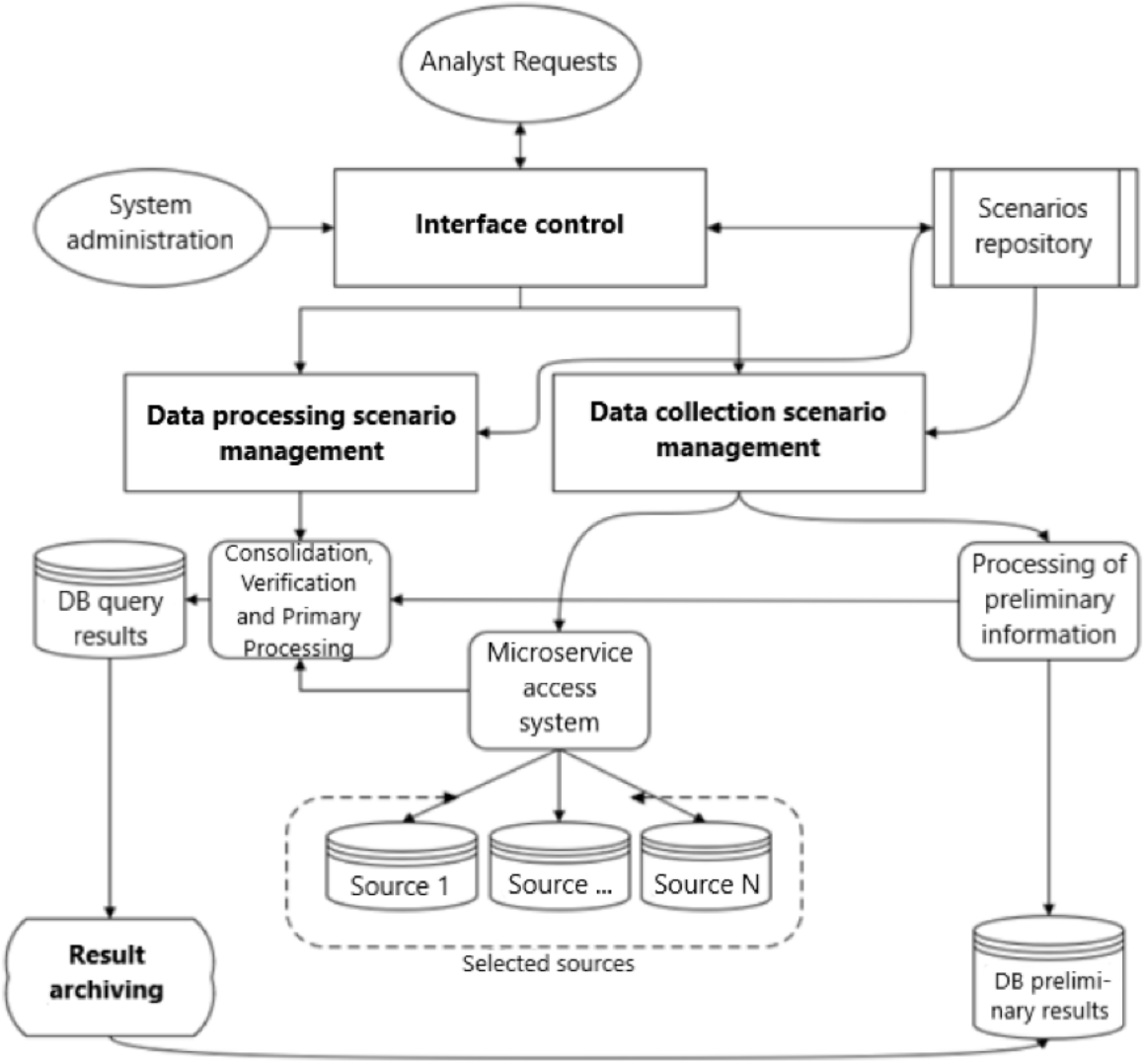

Fig. 1. Functional structure of the system

The interaction of software services in the system and the organization of individual information flows is based on an adaptive event-oriented architecture. This makes it possible to maintain high efficiency and throughput of the system for processing requests with a large number of users and processing complex requests, both from the point of view of large volumes of data and complex types of processing and analysis of this information.

By event, we mean a structured message of a fixed structure, which will be sent to a certain section in the message broker, with addresses for all services that are subscribed to update the message queue for a certain section.

With this approach to building the architecture, it is possible to obtain results in full in accordance with the tasks set by the analyst. No matter how complex and voluminous the request, which means the overall load on the system, ultimately the analyst (user) will receive the correct data that was specified in the request [6].

The main advantages of the presented adaptive event-oriented architecture are as follows:

– the ability to process streaming information in real time;
– possibility of parallel data processing;
– system scalability;
– speed of response to the analyst's request;
– the possibility of adaptation to the quantitative assessment of the relevance of information obtained during the execution of the analyst's requests.

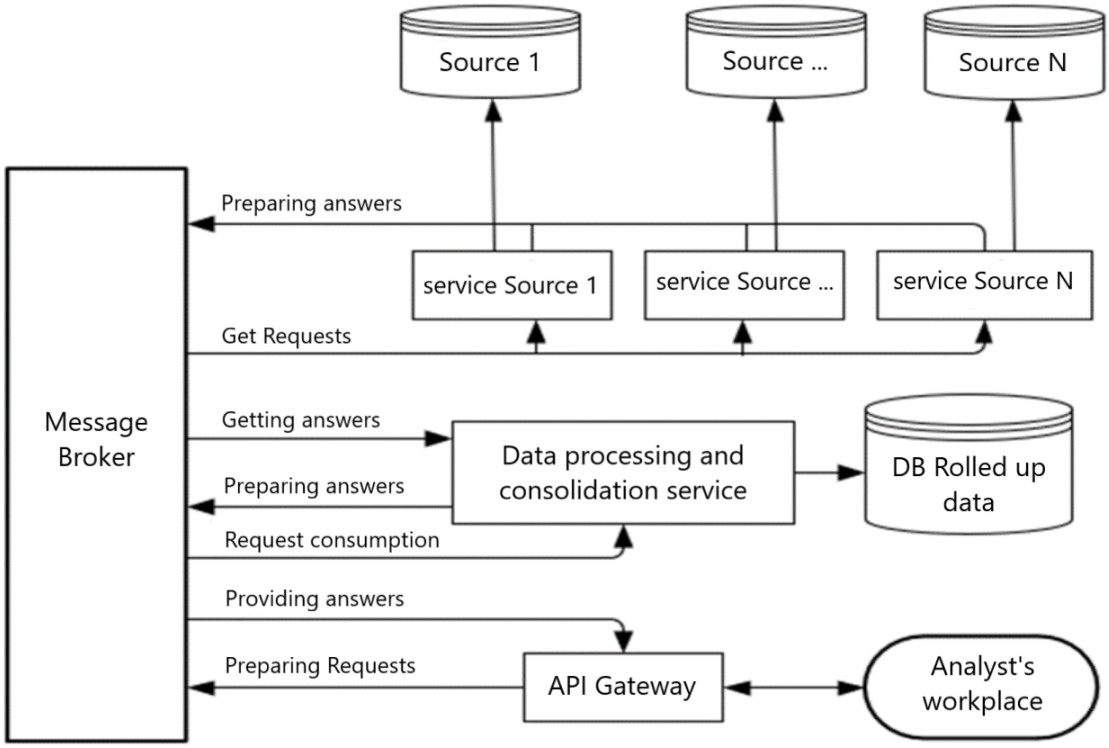

Fig. 2. Architecture of an adaptive event-oriented system

The architecture of the adaptive event-oriented system (Figure 2) is built according to the characteristics of data sources and methods of information processing, which is determined by the possible needs of user-analysts. Features and quantitative composition of data sources do not prevent partial implementation of the system, and the structure of the system implementation allows to increase the capabilities of the system as needs grow. The number and features of data sources affect only the implementation of individual applications aimed at working with individual sources. The implementation of separate services for processing results can also be carried out as needed.

The adaptive features of the system when collecting information from external sources are based on the processing and analysis of information about the relative relevance of the results obtained from the collection of information from external sources (usually bibliographic databases). Such information makes it possible to adjust the number of requests to these databases proportional to the relative stochastic assessment [5] of relevance, which thus significantly improves the efficiency of collecting such information according to requests. The implementation of adaptive event-oriented management of the collection of information from external sources, such as bibliographic databases, is based on the use of a linear stochastic automaton as a method that affects the formation of beats that implement chains of requests to data sources [6].

From a formal point of view, it is most convenient to describe the event handler as a finite stochastic automaton [7]. A state machine is described as a set of admissible states and a set of admissible transitions between them, that is, it is a directed graph.

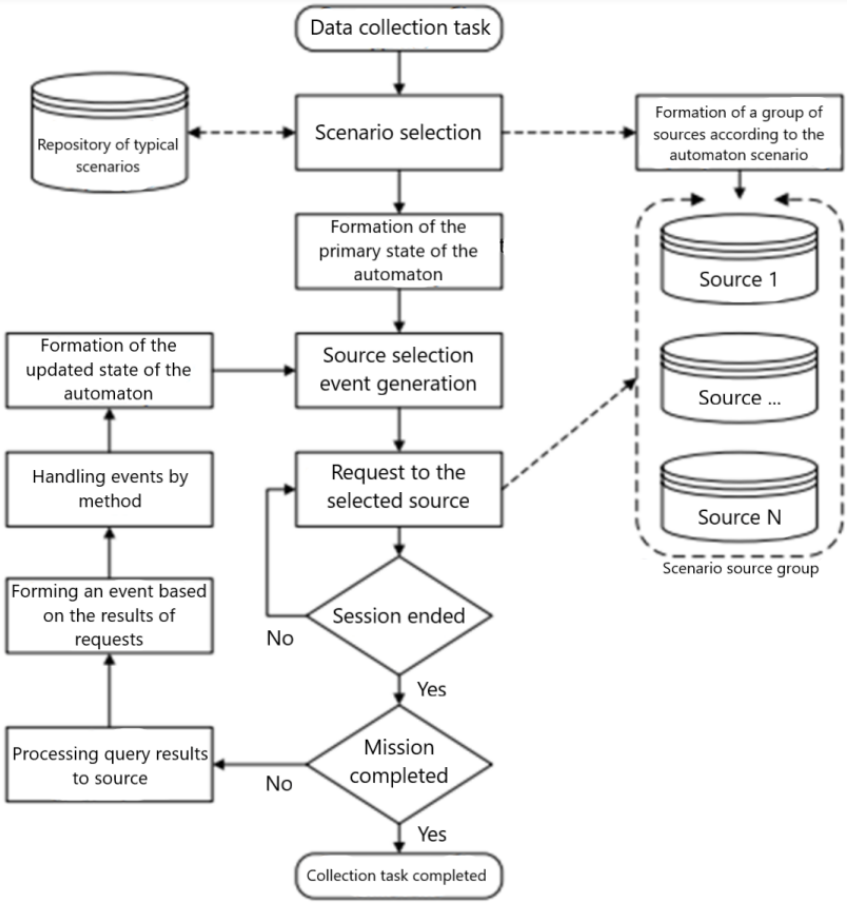

Fig. 3. Algorithm of using a stochastic automaton

In case of an event, the system calls the method of the object associated with this event. The method handles the event (perhaps generating events for other objects) and exits. If the methods of all handlers perform their operations quickly and without the use of blocking system calls, it is possible to ensure a high speed of response to events even within the framework of a single-running program [8]. If events arrive faster than the handlers can process them, the system queues them up.

4. Conclusions

The novelty of the product consists in the creation of an adaptive software system for collecting and processing consolidated information, which accumulates data on the international activities of individuals and legal entities with the possibility of flexible expansion of sources of information collection depending on the needs of users.

Among the users of such systems can be various scientific institutions, institutes, universities, research organizations, government agencies and other organizations that finance scientific research and individual scientists. This information and analytics allow scientific institutions to monitor their position in the world ranking, state institutions to direct the necessary attention to the level of research activity in the extremes, and investors to create a complete investment picture regarding the relevance of funding, for individual scientists to determine the distribution and citation of publications by global sources

The main goal of the system is to provide users with the opportunity to receive up-to-date analytical information on indicators of international activity from many selected sources.

The main tasks that the system solves:

- collection of data from sources of different structure and composition, primarily bibliographic and scientometric databases,
- aggregation of received data at the request of the user,
– deletion of erroneous data and duplicates of already received data.

Since the system is focused on a variable number of sources, which will differ in data structure and interaction interfaces, the main requirement is a flexible architecture with the possibility of constant modification with minimal costs for including new sources in the system, excluding old and unused ones. The most needed functions and modification of the still relevant existing ones.

The developed system makes it possible to collect information about the international activities of individual scientists or institutions from sources that are heterogeneous in terms of storage form and composition.

The composition of information sources in the system can be expanded without interfering with the operation of other system components. Each individual processor of information from the sources can provide only partial data, which is supplemented at the expense of others. The system adapts to the user's request in such a way that data is taken only from those sources that meet the requirements of the request and can satisfy them.

Also, the system is characterized by adaptability to the load and high throughput, which is achieved due to the use of modern approaches in the organization of the software system architecture.

The system provides users with a single point of entry for information requests and analytical information about the international activities of both individuals and institutions.

5. References.

1. Novikov DA Cybernetics: Navigator. History of cybernetics, contemporary state, prospects of development. - M.: 2016. - 160 p.

2. Ford Neal, Parsons Rebecca, Kua Patrick. Building Evolutionary Architectures: Support Constant Change. - O'Reilly Media, 2017. - 332 p

3. Chris Richardson, Microservices. Development and refactoring patterns. S.-P.: Peter, 2019. —544 p.

4. Adam Belnar. Building Event-Driven Microservices: Leveraging Organizational Data at Scale. USA, O'Reilly Media, 2020, 324 p.

5. Koval, OV, Kuzminykh, VO, Husyeva, II, Beibei, X., Shiwei, Z. Improving the Efficiency of Typical Scenarios of Analytical Activities. CEUR Workshop Proceedings, ISSN: 1613-0073 - Vol-3241,-2021- pp. 123-132

6. Mouett E. Using Docker / trans. with English AV Snastyna; scientific ed. AA Markelov. - M.: DMK Press, 2017. - 354 p.

7. Kuzminykh, V., Koval, O., Voronko, M. Standard Analytic Activity Scenarios Optimization based on Subject Area Analysis // CEUR Workshop Proceedings. ISSN: 1613-0073 – Vol -2577, - 2019, pp. 37-46.

8. Zhurakovskyi B., Toliupa S., Druzhynin V., Bondarchuk A., Stepanov, M. Calculation of quality indicators of the future multiservice network. in future intent-based networking: On the QoS robust and energy efficient heterogeneous software defined networks , - 2022, pp. 197-209.

9. Pethuru Raj; Jeeva S. Chelladhurai; Vinod Singh. Learning Docker. Packt Publishing, 2015. - 240 p.