This book is devoted to the problem of automation and integration of information processes in science and education. The history of the development of modern science and technology is associated with multidisciplinarity in scientific research and creation of research teams working in large projects. However, the current state of effectiveness of this work and integration does not fit current requirements. Therefore new approaches are needed for solving the above mentioned problem. In this work the analysis of the existing experience of using the distributed registry technology in various fields is conducted, the main features of the technology are considered for making decisions about its application in scientific informational activities, specific examples of its use and efficiency analysis are provided. The semantic core methodology as a universal tool for classification of unstructured information and qualification comparative analysis methodology for achieving a more rational use of human capital are presented. The book should be useful for students in Information Technology, scientists and graduate students in the field of Knowledge Management, scientific project managers.

Alina A. Ryazanova, dipl. in Regional Science and International Relations. Deputy Director of International Cooperation in the CCCDA (Center for Cryptocurrencies and Digital Assets). Postgraduate in Information Processes and Systems.

978-620-0-30195-6

Alina A. Ryazanova

Modern Technologies for Scientific Work

Approaches to the Information Processes Integration in Science and Practice

LAP LAMBERT Academic Publishing

Alina A. Ryazanova

Modern Technologies for Scientific Work

Alina A. Ryazanova

Modern Technologies for Scientific Work

Approaches to the Information Processes Integration in Science and Practice

LAP LAMBERT Academic Publishing

Imprint
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher:
LAP LAMBERT Academic Publishing
is a trademark of
International Book Market Service Ltd., member of OmniScriptum Publishing Group
17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page
ISBN: 978-620-0-30195-6

CONTENTS

CONCEPTUAL INTRODUCTION 3
PART ONE 6
DISTRIBUTED REGISTRY TECHNOLOGY. PRACTICE AND TRENDS 6
1. Global trends to controlled decentralized systems 6
2. Fundamentals of distributed registry technology 8
3. Advantages and disadvantages of using for cryptocurrencies 11
4. Categories 12
5. Areas of application 14
6. Main features of systems based on distributed registry technology 18
7. Summary 22
PART TWO 23
THE DISTRIBUTED REGISTRY TECHNOLOGY IN SCIENTIFIC-INFORMATION AND SCIENTIFIC-ORGANIZATIONAL ACTIVITIES 23
1. Basic aspects of using the distributed registry technology 23
2. Distributed registry for scientific information support 25
3. Distributed registry for voting and other feedback procedures 28
4.The cost efficiency of the distributed registry technology 31
5. Summary 32
PART THREE 34
A SEMANTIC KERNEL AS A UNIVERSAL INSTRUMENT FOR CLASSIFICATION AND SYSTEMATIZATION OF UNSTRUCTURED INFORMATION 34
1. Automated analysis in the HR with the help of semantic algorythms 34
2. A methodology for the formation of databases based on semantic analysis 36
3. Semantic analysis methodology application examples 39
4. Summary 44
PART FOUR 45
METHOD OF QUALIFICATION COMPARATIVE ANALYSIS OF REQUIREMENTS TO PROFESSIONAL SKILLS 45
1. Main definitions 45
2. Description of qualification comparative analysis technology 46
3. Application example: QCA in the specialty programmer 1C 48
4. Application example: QCA in the specialty systems analyst 54
5. Summary 59
COMMON CONCLUSION 60
REFERENCES 62

CONCEPTUAL INTRODUCTION

The history of science and technology in modern age, primarily since the mid-19th century, is associated with the development of multidisciplinarity in scientific and scientific-practical research, the creation of teams of scientists engaged with the same task within large and global projects.

It is historically worth mentioning first of all the “Manhattan Project”, targeted towards the creation of atomic weapons in the USA in the mid 40s of the 20th century, when thousands of scientists were first involved in the study and the need appeared to coordinate their efforts and to organize their work, first of all the by means of supplying them with important scientific materials and analyzing the results of their numerous studies.

At that time, it was not possible to solve this problem with any substantial success, and not only the project participants wrote about this fact in their memoirs (Richard Feyman), but even the project manager General Groves.

This was due not only to the lack of storage and processing tools for large amounts of information for organizing the work of researchers in a strict sense (computer technology was used in the project mainly for modeling of physical processes), but also to the lack of a methodology for organizing the work of large scientific teams.

In modern global projects, for example, mapping on the human genome and genome of other biological species, developing therapy against oncology diseases, the teamwork of scientists is presented more widely than before, but its forms nevertheless have not reached a qualitatively new level.

Dialectics teaches us, in accordance with the law of the passage of quantitative changes into qualitative changes, that with the increasing number of scientists and increasing volume of research data, a qualitative leap in the results of scientific activity should occur. One of the conditions for this breakthrough will be real integration within scientific activity, which will lead to a qualitative leap. In the meantime, this jump does not occur, which suggests that there are problems in integration.

The problems associated with the integration of scientific processes of researchers in large projects can be divided into several classes. First of all, these are technical problems associated with different linguistic environments and different levels of knowledge of the languages used for scientific communication, in addition, not all researchers have equally convenient access to both the scientific and information infrastructure. Further, this is the use of various software products by researchers, ranging from text editors with incompatible file storage formats to various email systems and messaging systems. The third group icludes problems of methodology, related both to the initially different approaches of researchers to the subject of research, and to different formats of the presentation of scientific material, for example, a different structure of articles and preprints, as well as different national systems for enumeration and recording of scientific results. The fourth group of problems is the reliability of the results and their correct time sequence, which is important for establishing scientific priorities and determining common points for the continuation and development of research in a given subject area. The fifth group of problems is related to the establishing “feedbacks” in scientific activity - updating the knowledge and skills of scientists and professionals involved in the technical support of scientific research, in accordance with the current level of both scientific knowledge, approaches and methods, and with skills for their implementation. This connects the task of integrating scientific processes tightly with the education and training of spesialists of high qualification.

Thus, in the modern scientific activity, it is necessary to speak generally about the integration of its processes.

Integration of processes - ensuring the universal interaction of subjects (for example, common data transfer interfaces between subjects) and providing a single format of objects used by subjects united in a process.

From the point of view of this monograph, the integration of processes is, from a practical point of view, the use of a distributed registry for data transfer and storage (as a common interface) and the use of semantic transformations as a tool for interaction between subjects (programs and program complexes). In this case, we understand the subjects and objects as elements of computer systems within the subject-object model - the subjects are active entities controlled by man, programs and program complexes, and objects are the data processed by these programs.

The statement of tasks to a large extent determines the structure of the monograph. The first part is devoted to the analysis of the experience gained in applying the distributed registry technology, its positive and negative effects, and the practice of using them in various fields.

The second part discusses the main properties of distributed registry technology for making informed decisions about selection and use of it in order to ensure scientific and information activities, provides specific examples of its use and analysis of effectiveness.

The third part of the monograph is devoted to the methodology for distinguishing the semantic core as a universal tool for classifying and systematizing unstructured information.

The fourth part discusses the methodology of qualification comparative analysis and its application in order to achieve a more rational use of human capital by improving the quality of training of qualified specialists and reducing education costs by eliminating obsolete disciplines from educational programs and correcting outdated disciplines. Concrete examples of the application of the methodology and recommendations for the correction of educational programs are given.

PART ONE

DISTRIBUTED REGISTRY TECHNOLOGY. PRACTICE AND TRENDS

1. Global trends to controlled decentralized systems

The level of development of science and education in the field of information technologies today not only does not meet the challenges and requirements of modern society, but also does not significantly correspond to its prospects and trends. This is due to a whole set of factors - starting from a stable technological lag and ending with the lack of awareness of the fact that the modern information society, including all information processes taking place in it, began to acquire a quantum-mechanical nature, therefore, is determined in accordance with the Heisenberg uncertainty principle. In application to information systems and processes, it can be formulated as the active influence of information exchange participants on the information processes.

We see this situation explicitly when decentralized systems are established, which function primarily on the base of distributed registries. The influence of initially few participants creates an avalanche-like self-regulatory process, covering all areas of activity and gradually involving the largest players who not only regulate and direct the processes, but also themselves become hostages of the processes created and launched by them. A striking example is the IBM project - HyperLedger Fabric.

This can be called a factor of involuntary reflexive influence of the system participants on the processes occurring in it, which significantly complicates the assessment and forecasting the development of modern information technologies, because this requires the inclusion of new aspects in the analysis, since the system participants also represent systemic integrity, having their own structures and processes in interaction with other system integrity. It is not possible to adequately assess all the prospects for the transition to decentralized systems, but the accumulated empirical experience makes it possible to identify the tendency towards a paradigm shift from a centralized model to a controlled decentralized one.

According to the Forum on the World Economy (2017), the fundamental technology for the Fourth Industrial Revolution will be the distributed registry technology, and in eleven years the share of transactions carried out through blockchain systems will be 10% of global GDP. But, despite the many projects and practical initiatives in the blockchain field, as well as growing public attention, very few scientific papers have been published on this topic that is extremely relevant and requires a systematic approach.

The distributed registry technology has already established itself in the field of finance, as well as the technical and possible legal aspects of its application have been studied, but there are still almost no systematic studies. In existing publications, the authors, as a rule, limit themselves to the study of particular cases of application, refusing system categories of a higher level.

Nevertthereless in the field of cryptocurrency and distributed registries, scientists have already proposed a number of technologies that are inaccessible for the largest players and realize a lot of expectations of both process participants and potential regulators - these are isolated cryptocurrency networks, authorized and reliably protected cryptocurrencies, the development of ideas and technologies of distributed registries (blockchains) in a two-dimensional block matrix.

These positive trends need organizational support and institutionalization.

All this testifies to the fact that it is extremely important for modern science to study and analyze accumulated experience in order to present the prospects and potential applications of the distributed registry technology. In the course of the research, a comprehensive analysis of the technical and system bases of the technology, structural chances and risks associated with its use is made, various models are created, the principles of operation of blockchain systems are described. At the same time, the blockchain is also considered as an information infrastructure solution.

2. Fundamentals of distributed registry technology

Distributed registries, in particular, the blockchain are defined as the type of database in which records are grouped into blocks. These blocks are connected in an unbroken chain in chronological order, which is provided by the hash value of each subsequent block, dependent on the hash value of the previous block. Each block contains entries created from the time it was added [2]. In the case of cryptocurrency, the bitcoin block includes the history of all completed transactions since it was added to the network. The principle of forming records in blocks distinguishes the blockchain from other distributed registries in which records are made continuously.

A feature of the blockchain cryptocurrency bitcoin system as the first largest decentralized system is that currency management is performed through a distributed computer network without the involvement of a central authority [1]. Therefore, it is the Bitcoin blockchain technology underlying the cryptocurrency that is actually considered as an innovative technology that, due to its properties, can significantly transform society. At the same time, transformations will take place far beyond the area of digital currencies.

Appropriate distributed registry management systems are described as consensus systems, since consensus algorithms ensure consistency between network nodes. Such systems are based on cryptography and Peer-to-Peer (P2P) principles to ensure the consensus of the entire network.

From the above definitions it can be seen that blockchain systems are distributed systems. Distributed systems can be characterized by several properties. First, the nodes of such systems interact and synchronize with each other. Secondly, the failure of a separate node does not adversely affect the operation of the system as a whole. In addition, each network node contains the information of the entire system, so the failure of one or several nodes can not lead to complete or partial data loss. In systems based on distributed registry technology, the entire chain of blocks is stored in each node [2].

Among the mentioned principles of "Peer-to-Peer" are the following. To provide data and network services, participants have hardware resources. In addition, there is a direct exchange between nodes, that is, there is no central node for coordination between network nodes. All this necessitates the use of cryptography for block chains. Both cryptographic methods and consensus algorithms with which network nodes coordinate the state of the system are the basis of the blockchain technology.

The principles of the blockchain technology can be considered on the example of the Bitcoin payment system, which is its first practical application [3].

Cryptocurrencies already existed for 20 years before the advent of Bitcoin. However, their significant drawback was the presence of an intermediary in payment transactions, so the first attempts were doomed to failure [4]. This drawback was eliminated with the beginning of the use of blockchain technology, since cryptocurrencies based on blockchain technology are digital currencies, in the circulation and management of which there is no intermediary. In this case, the decentralized peer-to-peer computer network performs and verifies transactions [5].

Bitcoin blockchain is a registry of all transactions conducted, formed in chronological order. A copy of this decentralized registry is kept on each member of the network. Cryptographic algorithms and decentralized management allow secure transactions. The functioning of the Bitcoin blockchain system is based on two principles of cryptography. These are public key cryptography and digital signatures, as well as cryptographic hash functions.

The concept of public-key cryptography was developed in 1976 [6]. In this case, the algorithm involves the generation of a pair of mathematically related keys - public and private keys. The key pair allows you to encrypt information so that it is available only to two users. For this, the sender supplies the message with a private key, which is known only to him, and sends the signed message to the recipient. Using the public key, the recipient of the document establishes the authorship of the document and the immutability of the document after signing.

Digital signature allows you to achieve three goals. Only the sender knows the secret key, so he can prove his authorship. Also, the sender can not deny that he signed the message. In addition, due to asymmetric encryption, changes do not go unnoticed, which ensures the integrity of the block content. Bitcoin blockchain uses a cryptographic hash function. A hashing algorithm is an algorithm that converts a string of any length into a string of fixed length (hash function).

The hash function is a deterministic function, i.e., the same input data will always have the same hash function, and any change in the input data leads to a change in the hash value. A special feature of cryptographic hash functions is the impossibility of both recovering encrypted data by a hash value, and finding other data that have the same hash value.

As mentioned above, one of the main achievements in the Bitcoin system is the consensus mechanism, which allows you to take a new block into the network. All network nodes reach a consensus on the adoption of a new block by means of a proof-of-work algorithm (proof of work), which is a set of requirements for complex computer calculations that must be carried out to create a block and add it to the blockchain. The objectives of the algorithm are to verify the transaction, which allows you to avoid the so-called double spending, and create a new Bitcoin unit.

The key feature is asymmetry: the calculations should be moderately difficult for the participant, but simple enough for the network as a whole. This is achieved using cryptography algorithms. Each member of the network is trying to find a solution to the problem first; at the same time, it can actually be found only by using the direct enumeration method, therefore, a successful solution requires many attempts. The newly generated block is a reward for time-consuming computations. In this case, transactions are randomly grouped into a block, and network participants (nodes) confirm the legitimacy of transactions in the block. Transactions are added to the block, which is now available to each member of the system.

The complexity of creating blocks in the Bitcoin network is due to the requirement of a high distributed computing power of the network. The higher the power, the greater the number of calculations that must be carried out to create a new unit [1]. The method also increases the cost of creating a block, encouraging participants to increase the efficiency of their systems in order to maintain a positive trend. The Proof-of-Work algorithm is used not only in the Bitcoin system, but also in many other cryptocurrency systems based on the blockchain. In each case, the specific features of Proof-of-Work may differ slightly, since they are created individually for each blockchain.

3. Advantages and disadvantages of using for cryptocurrencies

Note the following advantages and disadvantages of cryptocurrency compared to national currencies.

Benefits include:

• divisibility into sufficiently small units;

• an increased level of privacy, since there is no link to a bank account and no identification is required;

• minimal transaction costs and time costs due to the absence of intermediaries;

• The impossibility of falsifying a cryptomonet, which follows from the properties of the blockchain;

• Higher level of reliability due to the absence of intermediaries;

Among the shortcomings can be identified:

• significant fluctuations in the exchange rate of cryptocurrency; however, cryptocurrency volatility tends to decrease;

• irreversible nature of transactions;

• anonymity of illegal operations;

• possible hacker attacks, as digital assets are in electronic online wallets.

4. Categories

Since the blockchain technology is at the initial stage of its development, stable classifications by scopes began to appear only in recent years. The first classifications emerged as a result of the movement from the development of blockchain technology in the field of digital currencies to other areas of application. In this case, three phases are often distinguished [1]. The first phase includes currencies and related financial services applications. The second phase includes the following applications in economics and finance, such as, for example, smart contracts, which are described below. In the third phase, applications are developed that go beyond finance and markets, for example, blockchain applications in the public sector.

A significant contribution to the development of blockchain categorization was made by William Mugayar [7].

According to Mughayar's categorization, the blockchain technology's application environment includes infrastructure and platforms, middleware, applications and related services. Top-level categories, in turn, can be divided into several subcategories. In this case, the division into categories is based on the functional characteristics.

1. Infrastructure and platforms

An IT infrastructure is a combination of certain continuously used IT resources that form the basis for applying a technology. These IT resources include both hardware resources and operating systems, networks and information technology, as well as main data and data processing applications. In relation to the blockchain, this category includes, in particular, the blockchain system itself, multiplatforms for the development and implementation of various applications that are based on blockchain systems, as well as specialized hardware for the implementation of various algorithms.

Multiplatformism provides the opportunity to develop innovative derivative financial instruments that can be used in an economic or social context [1]. Such platforms are applications related to a specific task and are designed for the development and implementation of decentralized blockchain applications in blockchain technology.

2. Middleware is universal services that are the link between platforms and applications. They fulfill the requirements of a large number of applications. For example, the middleware includes smart contracts implemented on multiplatforms cryptocurrencies, or application programming interfaces for implementing applications in blockchain systems.

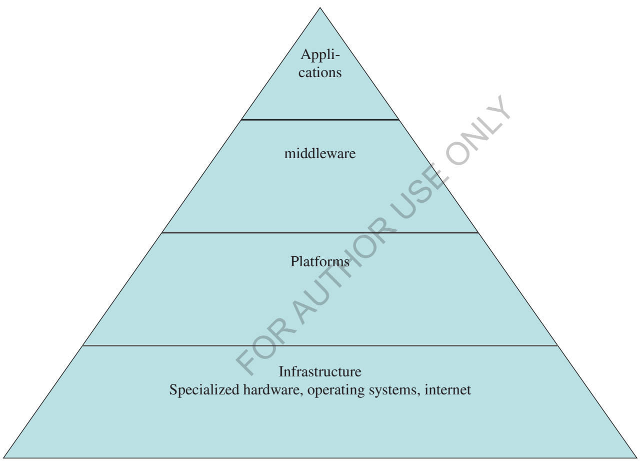
3. Applications are executed on the basis of the existing infrastructure and middleware. While the concept of smart contracts is related to middleware services, specialized private implementations of smart contracts are applications. An example of an application is the registration of digital and material objects in the blockchain system. In this case, the system generates digital documents containing the identification data of the object and the owner and stores them in encrypted form, i.e. The blockchain system does not store documents, but hash functions from the block contents.

The problem that negatively affects the implementation of such applications is the limited scalability of open blockchain systems, which limits their use in areas with a large number of objects to be taken into account. In addition, the exact identification of the object is difficult, therefore the inclusion of an object in the blockchain requires the participation of a trusted party.

4. Ancillary services

In addition to the categories described above, ancillary services are presented as the fourth category in the application environment of the blockchain technology. This category includes data providers, specialized media or industry investors. These services are provided, for example, by the website coingecko.com, which provides an overview, a comparative analysis, market capitalization indicators and a forecast for the development of various cryptocurrencies.

Figure 1 - Blockchain categories



5. Areas of application

The applications of the blockchain technology are not limited to the financial industry. Technology is becoming widespread also in many other areas of human activity.

The blockchain serves as a means of implementing smart contracts (transaction protocols responsible for executing contract terms). Information from an external source is used as input data, which triggers the execution of the protocol, based on the rules established in the smart contract.

The relevant contract details are recorded in real time and stored in the blockchain chain at a specific address. Smart contracts are a means of automating the actions of the parties to the contract. At the same time, contracts using appropriate algorithms can be executed, verified, terminated, and transfer of funds, transfer of rights, as well as rights management, are carried out transparently, safely, and blockchains remain unchanged in the chain. The inability to reverse the action under certain conditions can be regarded as an advantage. Autonomous execution of the contract excludes attacks and interference from outside, but it does not exclude the possible dependence of its execution on the occurrence of external events.

Even if the parties to the contract do not trust each other, smart contracts can ensure fair exchange between them without the participation of a trusted intermediary.

However, the use of smart contracts is also associated with certain risks, such as the unsettled relationship with contractual law and consumer protection and the associated conditions of legal liability, since contracts are executed through the program, rather than by traditional methods. In addition, there may be a shortage of specialists for the implementation of smart contract technology, which may negate the benefits from the use of technology across whole industries. Informally, the essence of the problem lies in the fact that while the smart contract operates with data and other elements inside the blockchain platform (for example, we are talking only about operations with cryptocurrency), we can talk about a certain security. However, if you try to expand the scope of distributed registries, the inevitable interaction with the external environment leads to the emergence of security threats of various kinds.

Formally, the statement about the insecurity of interaction with external data sources was formulated in [8].

Distributed registries are primarily used in the field of data management for recording data [9]. Such records have time stamps and are stored in distributed registries, which reduces the cost and complexity of managing them.

The distributed registries technology is also used in solutions designed to identify and confirm access rights [9].

The principle of identity in the blockchain system can provide users with greater control over access to their personal data and managing their openness to others.

The blockchain allows you to confirm and fix the right of authorship and scientific priority with the possibility of their publication after a specified period of time [10]. Using unique identifiers and digital certificates to confirm authorship and authenticity, the author can receive and retain the priority of his invention. In addition, it is also possible to create a mechanism for transferring ownership from the right holder to the buyer.

The blockchain technology is used to carry out transactions and store any form of money, goods or raw materials [2]. Legal entities and individuals can transfer funds to their account in the system using bank transfer, debit, credit cards, or from their bitcoin wallets.

With regard to the energy sector [11], the technology allows real-time measurement of the level of generation and consumption of electricity and other types of energy resources.

The blockchain technology with built-in cryptography can be used for anonymous voting [10], which guarantees the accuracy and reliability of the results, as well as the impossibility of their fraud.

The blockchain can be used to increase the transparency and integrity of political systems [9]. In particular, there is a whole international virtual nation in which there are citizens, ambassadors, partners and offices around the world. Anyone can join it without any restrictions.

As for solutions for effective management within organizations, specialized services are also created for this purpose, which automate the process of managing a company.

There are also blockchain platforms in the field of the Internet of Things [2] aimed at improving consumption practices. In this case, the system stores the identification data of physical objects with embedded microchips. This allows you to create digital identifiers that are safe and compatible with many other systems, which opens up opportunities for new mechanisms of interaction with the consumer, based on its proximity to the subject.

There are a number of applications for large-scale smart management of industrial systems and equipment. The basis of the development are the principles of decentralization, cryptographic protection and autonomy.

One of the first countries to use blockchain technology since 2013 for storing data in e-government databases is Estonia. At the same time, the prerequisites for using the technology were created from the beginning of the 2000s, when a digital signature and a decentralized data storage system ("X-Road") were officially introduced, in which all administrative-territorial units have access to all data. The decentralized structure and digital signature laid the foundation for the introduction of the blockchain technology into the Estonian e-government.

The Estonian government uses a Keyless Signature Infrastructure (KSI) based on the Guardtime blockchain to verify the authenticity of data in electronic registries. The data is stored in a distributed registry, which ensures their integrity.

Estonia offers its citizens hundreds of innovative e-services, including electronic filing of tax returns, issuing testaments, applying for state subsidies, and obtaining medical prescriptions.

Companies can send annual reports online, apply for licensing. Creation and registration in the state register of a legal entity can be done within a few minutes. Administration staff use the system to encrypt documents, securely exchange data, sign permissions. In the Cabinet of Ministers, a decentralized database is used to organize the meetings [12].

Thus, the experience of using the distributed registries technology, in particular the blockchain, allows us to distinguish the following properties of the information blockchain-infrastructure:

• openness: in open blockchain systems, any participant can implement their own product or application;

• heterogeneity: blockchain technology is applicable in various applications and areas. For different purposes, their blockchain structures are used;

• adaptability: further development of the technology implies its application for a variety of applications;

• recursiveness: certain recursive connections have developed around blockchains, in particular, in the form of so-called sidechains associated with parental blockchains, as well as other applications running on cryptocurrency platforms;

• uncontrollability: blockchain systems are by definition distributed systems, they do not have a trusted center, and the current status of the system is updated by participants using a consensus mechanism;

• distribution: the blockchain is contained in all nodes included in the network;

6. Main features of systems based on distributed registry technology

The structure of blockchain systems has certain advantages and disadvantages in comparison with other systems, which, however, also depend on the scope of the technology. The need for data encryption has led to the use of detailed access control procedures. Blockchain systems are considered anonymous and public since the address is represented by a public key, but in practice they can be called pseudo-anonymous (the public key serves as a pseudonym) [13]. Hashing is also an integral part of the functioning of blockchain systems and ensures that the data contained in the blocks is immutable.

The properties of a distributed peer-to-peer network in combination with a consensus mechanism for confirming the status of a network guarantee the fault tolerance of the system and the availability of data, since All nodes have a mechanism for verifying electronic signatures. The blockchain cryptocurrency using the consensus algorithm also solves the problem of double spending of funds. Each block passes a transaction legitimacy check, so trust between nodes of the network is not required.

The processes in the network run on the appropriate software code, which ensures technological integrity. Transaction history is open to all nodes, which predetermines a high level of system transparency.

In addition, blockchain tenology has the following advantages:

• Protecting large amounts of data with encryption and access control;

• the ability to collect and analyze large amounts of data from many enterprises;

• easier verification of data access points;

• automatic detection of vulnerabilities in supply chains, payment transactions and other business processes;

• reduction of IT infrastructure costs;

• reduction of expenses for internal and external financial transactions and management;

• the ability to promptly submit financial statements.

The blockchain technology also has significant drawbacks. In addition to the high energy costs of using the Proof-of-Work algorithm, there are limitations to the bandwidth and computing resources of some nodes, which is a critical issue with a significant increase in the number of transactions and participants [14]. The incompatibility of blockchain models leads to certain problems associated with the interaction with an open system, which are partially solved by the creation of sidechains.

In addition, there is a risk of loss or theft of the private key. In this case, making unauthorized changes can have serious consequences, since it is impossible to cancel a transaction. Also, in the case of a successful attack, various kinds of manipulations are possible, allowing the re-launch of the proof-of-work algorithm, starting with the first modified block, and therefore double spending of funds and creating an additional unit of cryptocurrency (for example, 51% attack, named after with the amount necessary for the successful implementation of the attack computing power).

Among the shortcomings should also indicate:

• insufficient scalability;

• low data rate;

• limited disk space;

• difficult management of access rights;

• the need for a high-speed Internet connection;

• complex integration with existing systems in the organization.

Summarizing, we can single out the following significant obstacles to the introduction of blockchain technology:

- Modern technology limits the scalability of blockchain systems and makes it impossible to minimize transaction risks (making entries in a distributed registry is irreversible);

- there is no legislative regulation of the use of smart contracts [14];

- the necessary standards for the application of technology have not been developed;

- not a high level of technology for the successful implementation of blockchain systems.

When creating blockchain systems, all sorts of factors that influence the success of the implementation should be taken into account, such as the possibility of creating a peering network, a transparent and reproducible history of actions. At the same time, the benefits of smart contracts for process automation can be fully manifested if each action of the process and its consequences have a clear description.

Despite the extremely optimistic views associated with the use of blockchain technology in various fields, it will take a long time to fully appreciate the effect of its use. Today, it can be argued that, due to its versatility, the technology deserves close attention and system research, followed by the creation of a blockchain laboratory for more detailed and systematic study of practical implementations.

In addition, it is advisable to create state services for regulating the circulation of digital assets and the use of distributed registries, whose functions include the coordination of standards (rules) in this area, monitoring the work of organizations in the use of digital assets, analytics in the field of cryptocurrency movement, bringing it to interested departments , supervising the work of distributed registry operators.

The main argument in favor of creating such a service is the need to address fundamentally new issues and problems arising from the circulation of digital assets and the use of distributed registries that cannot be localized within one or more departments, as well as the above-mentioned reasons for the methodological plan, which also impede the work of existing departments that are within the framework of established principles and decision-making technologies. Also, state-controlled activities of this type of activity will help prevent semi-legal operations and the overall imbalance in the field of finance.

In this regard, it is necessary to create a technological platform for digital transformation, which satisfies the following fundamental requirements:

- power of attorney (universal open structure and open source code) [4];

-use of national standards for cryptographic protection of information;

- versatility (the ability to use technology in various areas of activity: in the field of finance, smart contracts, to fix scientific priority and confirm copyright, to ensure the functioning of information systems in the public sector, etc.).

7. Summary

For modern science, it is extremely important to study and analyze accumulated experience in order to present the perspectives and potential areas of application of distributed registry technology. This article analyzes the technical and system fundamentals of the technology, structural chances and risks associated with its use, describes the principles of operation of blockchain systems. At the same time, the blockchain is also considered as an information infrastructure solution.

Blockchain as a technology of distributed data storage is able to stabilize information systems in the areas of public administration and science, but it can destabilize the sphere of finance and the economy as a whole. In this regard, it is necessary to create national technological platforms for digital transformation that satisfies the requirements described in this paper.

PART TWO

THE DISTRIBUTED REGISTRY TECHNOLOGY IN SCIENTIFIC-INFORMATION AND SCIENTIFIC-ORGANIZATIONAL ACTIVITIES

1. Basic aspects of using the distributed registry technology

The relevant and promising model of distributed data storage within a distributed registry (blockchain) [15] has been widely discussed in the theoretical and practical contexts.

The goal of this part is to explain the basic properties of the distributed registry technology to allow informed decisions on its selection and use in information activities in science.

As is obvious from the “second” name of the technology, as mentioned above a blockchain is a continuous chain of information-containing blocks organized in a certain sequence, which must primarily possess specific chain properties such as continuity and strength [15]. Continuity is defined as the property of positioning the blocks in a sequence, which is specified at the phase of blockchain creation, while strength represents the impossibility of replacing or removing any link from the chain.

What makes a blockchain advantageous for scientific and information activities? Currently, information is usually sent out via e-mail (newsletters). The receivers of scientific information cannot reach out to their source and provide feedback on whether the materials meet their scientific interests. As well, the information supplier cannot determine the parameters of information consumption (such as whether receivers read the material or what amount of time they spent on them) in an automated mode.

When the circular broadcast scheme is applied corporate network traffic is used in a suboptimal manner (a letter intended for many addressees within the corporate network is repeatedly duplicated and sent to the computers of the information consumers.

The distributed registry technology avoids traffic, which inevitably increases when information is sent out to a list of receivers, and can guarantee data protection from various intruders while providing data exchange without any intermediaries.

The structure of the blockchain link contains information as a control sum (hash value) about both the content of the link and its atoms and the structure and content of the preceding and succeeding links [1]. In this context, introducing any change to the atom of a link will inevitably affect their control sum when using reliable cryptographic algorithms. This in turn will require changing the control sum of both the preceding and succeeding links, i.e., the entire chain, since an exact match between data and the hash functions of the links leads to the strength, continuity, and hence, the security and reliability of the chain.

Thus, the blockchain technology provides a high degree of transparency of the completed operations and the ability to store and transfer data reliably with-out a risk of unauthorized modification by third par-ties and thus helps to increase the level of trust in scientific and information activities.

Being equally accessible to all the participants of information processes, the blockchain technology can provide the functionality of the feedback mechanism, which is a necessary condition for balance and effective interaction in the system of scientific and information support.

The blockchain as a system whole has a certain structure and its interrelations. Blocks or links in this system can be reasonably viewed as individual components consisting of so-called atoms, which, depending on their properties, can be of the following types [15]:

a boundary atom (for example, the start of a block) contains a reference to the end of the previous link;

a structure atom defines and describes the vector of identifiers of the link atoms;

a data atom contains a description of the data and is a passive element in the blockchain link;

a subject atom contains a description of the data and is an active component of the link.

2. Distributed registry for scientific information support

Let us look at the technology of a newsletter by including the information in the blockchain links (Table 1). Let us assume that a user named *Pi* is sent a document named *Fj*. The link also contains the function value *H*(*Fj*), which significantly depends on all symbols, letters, and fields in the document. This function is usually a hash function or a control sum. The link also has service information regarding the use of the document by the receiver (the fact of openingand reading, the time spent for reading, the usefulness of the materials to the receiver, etc.).

Table 1. The structure of the block for scientific information support

ID	Atom name	Content	Note
1	Boundary atom (start)	Reference to link Zi-1	
2	Structure atom	Description of the structure (user name field? Document field, control sum and service data for feedback	Required for the operation of softwaretools that process the block
3	Atom—open data	Pi	Name of the user for whom the document is intended
4	Atom—open	Fj	Document intended for the

	data		user
5	Hash atom	H(Fj)	Control sum
6	Atom—open data	Service information on whether the document has been read by the user	Fact of reading, time of reading, usefulness of the information for the user
7	Boundary atom (end)	Reference to the block Zi+1	

If there is information intended only for *Pi* with an identifier of *Xi*, atom-encrypted data are added to the blockchain. The following link structure can be used for a commercial newsletter that contains individual scientific materials (Table 2).

The scheme in Table 2 can also be expanded to focus on the scientific priority. In this case, the author of the publication, invention, or discovery can define a substantial workload for sweeping the *Xi* key in atom no. 4 (see Table 2) in order to publish the result at an appropriate time with a documented validation of the priority.

Table 2. The structure of the blockchain for a commercial newsletter with individual scientific materials

ID	Atom name	Content	Note
1	Boundary atom (start)	Reference to link Zi-1	
2	Structure atom	Description of the link structure (user name field?	Required for the operation of softwaretools that process the block

		Document field, control sum and service data for feedback	
3	Atom—open data	Pi	Name of the user for whom the document is intended
4	Atom—encrypted data	*E*(*Fj*, *Xi*)	Encrypted document intended for the user
5	Hash atom	H(Fj)	Control sum to check whether the decryption is correct
6	Atom—open data	Service information on whether the document has been read by the user	Fact of reading, time of reading, usefulness of the information for the user
7	Boundary atom (end)	Reference to the block Zi+1	

If the so-called “blind method” is required for reviewing the publication (within the *double-blind peer review* procedure) [16], where the authors (and some-times the editor as well) are not allowed to know the names of the reviewers and the reviewers are not allowed to know the names of the authors, the scientific field and the article text are included in the atom containing the open data and the author’s full names encrypted using the personal IDs of the authors (which is available only to the authors) are included in the atom that contains encrypted data. Information about the authors can only be disclosed by the authors themselves. Another constructive approach is to include the information in the atom link to which the hash value of the author's full name is written (the hash atom). It is impossible to recover the full name of the author via the hash function, while the authorship can be checked at any time by calculating the hash function of their full name.

3. Distributed registry for voting and other feedback procedures

Below is an example of forming a blockchain link for feedback procedures (including secret voting or approval of documents for a scientific project).

Let us assume that there is a voting procedure with *N* participants (*N* = 1, 2, …, *i*, … *N*). The result of the voting can be either binary "yes" or "no" or an arbi-trary string such as a candidate's full name or an opin ion on the scientific value of the document or the result of scientific research.

In this manner, *Ui* is a voting participant (Table 3); *Gi* is their vote or opinion (a text string in any case); *Xi* is a secret identifier (ID) created personally by *Ui* using a random number generator and is unknown to anyone; [*Xi*] is a secret ID truncated in such a way as to provide the defined workload sweeping (such as a day); and *Y* = *E*(*x*, *k*) is the algorithm of data encryption at the key *k*, where *y* is the encrypted data. Without knowing the key *k*, the data *x* cannot be recovered by a known *y*, which is a basic principle of cryptography. As an example, if the Russian Kuznechik of 2015 standard for data encryption is used as *E*, the workload of sweeping the [*Xi*] at the length of 19 bits on a house-hold computer takes approximately 30 s; when the length is extended by one bit it grows by 2 times.

The voting participant *Ui* will then form the following blockchain link *Zi* (Table 3).

When collecting the votes, the voting organizer forms and signs the links of blockchain *Zi* with their electronic signature, and then, during the tally of votes, performs sequential sweeping of keys [*Xi*] on the basis of the content of the fields (atoms) nos. 4 and 5(see Table 3), thus capturing the participant's vote as a result of decrypting atom no. 4.

Table 3. The structure of the blockchain link for a voting procedure

ID	Atom name	Content	Note
1	Boundary atom (start)	Reference to link Zi-1	
2	Structure atom	Description of the link structure (user name field? Document field, control sum and service data for feedback	Required for the operation of softwaretools that process the block
3	Atom—open data	Ui	Name of the voting participant
4	Atom—encrypted data	*E*(*Gi*, *Xi*)	Encrypted vote of the participant which can only be disclosed by the participant
5	Atom—encrypted data	*E*(*Gi*, [*Xi*])	Encrypted vote of the participant which can only be disclosed by the voting procedure organizer by sweeping the defined workload
6	Hash atom	*H*([*Xi*])	Information for the correct choice of the key that is being swept

7	Signature atom	Electronic signature of the voting procedure organizer under fields1–6 and 8	Information for fixing the block continuity
8	Boundary atom (end)	Reference to the block Zi+1	

This procedure eliminates the possibility of falsifying the voting data and reduces the pressure on the voter, as their vote is unknown until the tally of votes, as well as ensuring the sequence of voting and credible archiving of the voting results as blockchain links.

In order for the system of scientific and information support to function using a blockchain, the following groups of requirements can be defined:

structural: the requirements relate to the presence of any types of data to provide operation of the defined algorithms in the blockchain links. Structural requirements include the requirement of global continuity of a blockchain, i.e., the presence of a link or several links that describe the structure of the blockchain as a whole or its subset;

organizational: during the formation and processing of the atoms in a blockchain, national, certified, or recommended cryptography methods must be applied and the requirements associated with the national or institutional norms must be met;

technological: the requirements on reliability of storing the links (for example, during the use of the technology described in [17]). Storage of links in a blockchain must comply with the parameters of reliability of storage and accessibility of links established by the regulatory authorities of the respective industries. As well, the technological requirements must present the requirements on efficiency of the operations in a blockchain and the maximum volumes of data to store in them;

trust requirements: a blockchain must have a strictly defined structure, regulated technologies for working with its links, and an interface for operations with the links. In order to provide a high level of trust, all the applied interfaces must be available as the source codes.

4.The cost efficiency of the distributed registry technology

The blockchain technology can not only solve the problems of effective distribution of scientific materials, focusing on the scientific priority, voting, project management, but also generally improve the quality of scientific information and organizational activities, optimize the IT infrastructure of the projects, reduce the costs of owning scientific and information systems, and objectively monitor the effectiveness of work in scientific organizations.

Let us look at the efficiency rates of the technology. We will perform a cost efficiency analysis of using the blockchain technology in scientific and information activities for the following parameters:

- cost of information storage,
- traffic volume,
- architecture of information storage,
- saving on investments in an IT project.

Cost of information storage. Currently, the average cost of storing information in general-purpose data processing centers in Russia is approximately 30–40 rubles per gigabyte per month [18]. In corporate data processing centers, this number is 3–5 times higher. According to the assessments of experts from abroad, the price of storing 1 GB of data at a bandwidth of 30 GB per month is $1.51 [19]. Considering the more compact storage of information in a blockchain link, the volume of storage can be reduced by 2–2.5 times, which will lead to a corresponding reduction in the cost of storing information.

Traffic volume. Due to the optimized structure of data and the maximal optimization of the cryptographic algorithms, the volume of service traffic can be approximately 1.7 times lower. Moreover, interdepartmental use of the blockchain technology will allow preventing traffic duplication for e-mail newsletters and access to the database. According to expert assessments, this reduction is by approximately 4.8 times.

Architecture of information storage. The blockchain has become a "key–value" database [7]. Searching in such a database is only possible via a primary key and the volume of stored data is very limited. For applications that process scientific information this is clearly not enough. Thus, when developing applications on a blockchain, e.g., for *Ethereum* and *Masterchain*, the problem of storing and processing data is very acute.

The blockchain technology contains universal interfaces for forming and accessing data, which can be integrated to any application and provide operation of analytical systems and systems of scientific and information activities and minimize the corresponding costs at the phase of development and deployment. According to the expert assessments, the use of standardized interfaces reduces the cost of development, owner-ship, and maintenance by approximately 25–30%.

Saving on investments in an IT project. Universal interfaces that can be integrated into any application and provide cooperation of the created IT systems can minimize the costs at the phase of development and deployment.

5. Summary

The total cost advantage of the use of the blockchain technology in scientific and information activities can comprise at least half of the funds for forming, implementing, and supporting IT projects in the scientific environment. It can reduce the costs of providing various scientific and information services by several times while significantly increasing both the number of scientists who are provided with scientific and information support, as well as the promptness of providing information services to them via notably faster processing and transferring of data, while preserving the continuity and confidentiality of scientific materials.

Thus, from a practical point of view, the use of distributed ledger technologies as data transfer and storage platforms is a prerequisite for the quality integration of information processes to ensure information activities in science.

PART THREE

A SEMANTIC KERNEL AS A UNIVERSAL INSTRUMENT FOR CLASSIFICATION AND SYSTEMATIZATION OF UNSTRUCTURED INFORMATION

1. Automated analysis in the HR with the help of semantic algorythms

The development of science is not only the most important component of modern society, but also has an inverse effect on the development of society, that is, it contributes to the creation of effective human resource management tools, including those aimed at establishing continuous and objective automated (partially automated) monitoring and analysis of matching skills and competencies of professionals and therefore educational programs, to the existing requirements of the economy.

This part considers a technique and software complex for isolating the semantic kernel of a text array to ensure a high degree of accuracy in the selection of specialists for relevant vacancies, the requirements for which are specified in an arbitrary text form. Such selection makes it possible to compile a stable database of vacancies and a corresponding resume database. The availability of these two databases opens the path for further highly effective automated analysis of key skills, implemented on the basis of the semantic kernel.

From the point of view of human-resource management, the main efforts in the process of regulating the labor market at the current level are aimed at removing obstacles in the search for a workplace and an appropriate employee. There are two main reasons that qualified specialists do not find a suitable work-place. The first of these is related to the fact that the employee and employer describe the same range of skills and competencies using different terms and use different names for vacancies. The reverse situation is also possible, and even more common, where one term (for example, a technologist) describes fundamentally different specialties (for example, a food production technologist, a sewing equipment technologist, or an engineering communications technologist). In this case, it is necessary to use more sophisticated search systems than the search for the headers of ads, which is established in almost all modern employment sites.

Technological aspects of research in the field of human capital development are related to the means of interaction among the labor market, the education system, and the technological infrastructure, which makes it possible to find specialists who are suitable for a particular job, as well as to reveal systemic gaps in the training of specialists. The algorithm of interaction in the first approximation can be presented in the form of a “black box,” where the input contains the requirements of the employer for candidates for vacancies and characteristics of real employees and their resumes and the output is the conclusion about the compliance of qualifications of employees with the labor market requirements and, as a consequence, proposals for state or municipal bodies of professional education in the field of professional development or retraining of workers.

Automated data processing and the search for relevant information are greatly simplified through a single vocabulary of terms. However, a single updated dictionary of specialties, narrowly characterizing a certain type of activity, has yet to be created. To prepare a block of vacancies from the resume database most fully complying the advertisements, the methodology of multiple semantic text analysis is applicable. At the same time, it is of fundamental importance to automate the analysis procedure, preceded by studies of tests that are necessary for determining the level of preliminary manual processing of files.

2. A methodology for the formation of databases based on semantic analysis

Since the requirements of an employer for job candidates and the characteristics of real employees and their resumes are unstructured text data, the content of the black box can be processed using various types of semantic algorithms. These include text indexing, their comparison, identification of meaning, statistical analysis of texts, and searching in texts. In the first approximation, the task of finding the resume that is most suitable from the employer's point of view is to compare two unstructured texts.

The task of comparing the two texts in computer science belongs to classical ones. A sufficient number of semantic textual search methods have been developed based on the analysis of the frequency of occurrence of words, their correspondence to a certain topic, and their joint use and probabilistic distribution.

If the first word of the first text is selected and com-pared with all the words of the second text when it is found it can be stated that this word occurs in both the first and second text, while if it is not found, then this word is only in the first text. After the completion of the procedure, the second text contains the words found only in it. On average, the complexity of this procedure is the product of the lengths of texts in words by the average length of the text.

For the optimal solution of this problem, the apparatus of non-biased (ambiguous) mappings with the following properties is used.

Let W be a word of an arbitrary length L in some alphabet A. Consider the transformation $H(W) = h$, which maps words of arbitrary length to a word of fixed length.

This transformation must have the following property: for a random equiprobable choice of two words W_1 and W_2 in the alphabet A from the set of possible words, their corresponding words h_1 and h_2 must be different with high probability.

If the transformation H is a Shannon stirring transformation, then, as a rule, the length of the word h is used to estimate the probability.

Suppose that the length of the hash word is 3 bytes. Then, the conditional probability P (h_1 = h_2/W_1 is not equal to W_2) is estimated to be of the order of 2^{-24}, that is, 10^{-7} (taking the fact into account that $2^{10} = 1024$ is approximately $1000 = 10^3$).

Therefore, the following transformation can be applied to any text in any language: every single word of the text W that is longer than the length h can be replaced by the value (hash value) of its function H (hash function). For generality, the hash value can replace all words regardless of their length.

As a result, the text is converted into a sequence of binary numbers, let us call them hash words, each of which is of the length $|h|$, that is, the length of the hash value (in the example given, 3 bytes). The principal result of this transformation is that any constructions for comparison and searching become equal in length and there is no need to compare words of different lengths.

Further, for each text T_i, simultaneously with its transformation to hash words, a dictionary D_i is constructed consisting of non-repeated hash values and the corresponding words.

Dictionary D_i is a metric of the content of the text and makes it possible not only to optimize the search in the text (we are primarily looking for the words of the search query in the dictionary; if available, we search for them in the text, as modern search engines do), but also compare texts containing unstructured data between each other.

The complex for text indexing and analysis (CTIA) by Prof. A.Yu. Shcherbakov operates on the basis of the described algorithm [21]. The text indexing program m_ind, when launched in the format m_ind[.exe] file-name.ext, creates three files:

filename.csv: a list of words (in the Windows encoding) that occur in the indexed text (dictionary). The file consists of 35-byte records, of which 32 bytes are space-separated words, the ";" character, and two newline characters;

filename.lmd: index file;

filename.num: a file of double-byte values; the *i*th field is equal to the number of words with the *i*th entry number in the dictionary, found in the indexed text.

The text-comparison program tcmp when launched in the format tcmp [.exe] filename1.ext1 filename2.ext2 produces a set-theoretic comparison of the two texts specified in the arguments (the files filename1.ext1 and filename2.ext2 must be previously indexed by the m_ind program) and creates three files:

- onlyone.csv: words that occur only in the first text (filename1),
- onlytwo.csv: words that occur only in the second text (filename2),
- common.csv: words that occur in both filename1 and filename2.

The program estimates the similarity metric of texts in files filename1.ext1 and filename2.ext2. For this, three metrics are defined.

Let T1 be the number of words in the first text (file-name1.ext1), T2 be the number of words in the second text (filename2.ext2), O1 be the number of words in the file onlyone.csv, O2 be the number of words in the onlytwo.csv file, and C be the number of words in the common.csv file.

Then, C is the cardinality (the number of elements) of the intersection of the two sets filename1.ext1 and filename2.ext2. The similarity metric of texts should be equal to:

0, if C = 0, that is, if the intersection of the sets of two texts is empty;

1, if O1 = O2 = 0 and the texts are the same.

In this case, the equality T1 + T2 = O1 + O2 + 2C holds, which follows from the equalities T1 = O1 + C and T2 = O2 + C.

Then, we can define metric R1 = 1/2(C/T1 + C/T2). As experiments show, this metric will take maximum values when assessing the similarity of texts.

In addition, we can define a “natural” metric R2 = 2C/(T1 + T2), which, as is easily seen, takes the value 0 if the texts do not match (C = 0), and 1 if C = T1 = T2 in case of matching texts.

Finally, we can also define a third and minimal metric of the similarity of texts R3 = C/(O1 + O2 + C).

Metric R3 is necessary for cases when some words are deleted during comparison of texts, for example, words of short lengths (2 symbols: prepositions, con-junctions, and interjections) or other words at the choice of the analyst.

3. Semantic analysis methodology application examples

Let us illustrate this with concrete examples. For this we take one typical requirement for a vacancy and four resumes of applicants from the site hh.ru. For the analysis using the tcmp program, we have five files: base_t, which in an unstructured arbitrary form contains the requirements of the employer to the vacancy engineer–technologist in the field of oil production and refining; the resume engineer–technologist in the field of oil production and refining of the first applicant, file rt1; the resume engineer–technologist in the field of oil production and refining of the second applicant, rt2 file; as well as, for illustration, the technology files of two resumes: sewing manufacturing engineer–technologist, files rt3 and rt4. Here are the results of a pairwise comparison of all the resumes with the file base_t.

For the first applicant:

Min word length in COMMON => 0

Read pages…………….

Successful comparison! See onlyone,onlytwo and COMMON files

Files:

[rt1.txt]=398 words [base_t.txt]=171 words All=569

[onlyone]=347 [onlytwo]=120 [common]=51 All=569

Files metrics is correct

first Equal metric = 0.213193 [21%] –>Hihg Null-Equal metric = 0.179262 [17%] –>Medium second Equal metric = 0.098456 [9%] –>Down Medium = 0.163473 [16%]

For the second applicant:

Min word length in COMMON => 0

Read pages…………….

Successful comparison! See onlyone,onlytwo and COMMON filesFiles:[rt2.txt]=181 words [base_t.txt]=171 words All=352[onlyone]=158 [onlytwo]=148 [common]=23 All=352

Files metrics is correct

first Equal metric = 0.130787 [13%] –>Hihg Null-Equal metric = 0.130682 [13%] –>Medium second Equal metric = 0.069909 [6%] –>Down Medium = 0.110349 [11%]

For the third applicant:

Min word length in COMMON => 0

Read pages…………….

Successful comparison! See onlyone,onlytwo and COMMON files

Files:

[rt3.txt]=277 words [base_t.txt]=171 words All=448

[onlyone]=252 [onlytwo]=146 [common]=25 All=448

Files metrics is correct

first Equal metric = 0.118226 [11%] –>Hihg Null-Equal metric = 0.111607 [11%] –>Medium second Equal metric = 0.059102 [5%] –>Down Medium = 0.096215 [9%]

For the fourth applicant:

Min word length in COMMON => 0

Read pages…………….

Successful comparison! See onlyone,onlytwo and

COMMON filesFiles:[rt4.txt]=315 words

[base_t.txt]=171 words All=486[onlyone]=275 [onlytwo]=131 [common]=40

All=486

Files metrics is correct

first Equal metric = 0.180451 [18%] –>Hihg Null-Equal metric = 0.164609 [16%] –>Medium second Equal metric = 0.089686 [8%] –>Down Medium = 0.144771 [14%]

In these results of the program, the metric R1 occurs for the text first Equal metric, metric R2 occurs for Null-Equal metric, and metric R2 occurs for second Equal metric.

As is easy to see, the resume of the first applicant as much as possible corresponds to the requirements of the employer; the first metric takes the value of 21% (first Equal metric = 0.213193 (21%)). For the other applicants, the first metric takes the values 13, 11, and 18%, respectively.

It is not difficult to explain the fact that the resume of the second applicant is much less consistent with the requirements of the vacancy: his experience in the field of oil production and refining is less than 3 years, while the first applicant's experience is more than 18 years; it is much smaller in volume and contains fewer skills in the specialty.

An unexpectedly high metric of similarity was revealed in the comparison of the fourth resume sewing manufacturing engineer–technologist of with the basic vacancy (the length of the text of the resume is comparable to the length of the text of the first resume), which indicates the presence of a large number of words common to most engineer–technologist specialties in different fields of activities, such as production, technological, products, development, preparation, compilation, etc.

To improve the proposed methods, we perform the following. Consider, for example, the specialty engineer–designer of radio electronic equipment.

We combine all the requirements for the vacancy engineer–designer of radio electronic equipment in one text file and apply the m_ind transformation. As a result, we obtain a vocabulary of the vacancy that con-tains over 500 positions: basic words that describe the requirements for the applicant that occur at least once in the combined text. Next, we carry out manual expert processing of the text: we exclude prepositions and conjunctions, auxiliary words, and words that are of little importance for the specialty. If the first file is truncated to 231 words, we compare it with the resume of the radio electronic equipment engineer, sewing manufacturing engineer, and oil industry engineer– technologist (Table 1).

Table 4 Comparative analysis of vacancies and a group of resumes using the complex of text indexing and analysis (KTIA), truncated at up to 231 words

Vacancy	Resume	Average value from 3 metrics
Engineer–designer of radio electronic equipment	Engineer–designer of radio electronic equipment	0.101325 [10%]
Engineer–designer of radio electronic equipment	Sewing-manufacturing engineer	0.074569 [7%]
Engineer–designer of radio electronic equipment	Oil industry engineer–technologist	0.032651 [3%]

For the first applicant:

Min word length in COMMON => 0

Read pages…………….

Successfull comparison! See onlyone,onlytwo and COMMON files

Files:

[treb2.txt]=231 words [rezrea.txt]=235 words All=466

[onlyone]=203 [onlytwo]=207 [common]=28 All=466

Files metrics is correct

first Equal metric = 0.120181 [12%] –>Hihg Null-Equal metric = 0.120172 [12%] –>Medium second Equal metric = 0.063927 [6%] –>Down Medium = 0.101325 [10%]

For the second applicant:

Min word length in COMMON => 0

Read pages…………….

Successfull comparison! See onlyone,onlytwo and COMMON files

Files:

[treb2.txt]=231 words [rt4.txt]=315 words All=546

[onlyone]=207 [onlytwo]=291 [common]=24 All=546

Files metrics is correct

first Equal metric = 0.090043 [9%] –>Hihg Null-Equal metric = 0.087912 [8%] –>Medium second Equal metric = 0.045977 [4%] –>Down Medium = 0.074569 [7%]

For the third applicant:

Min word length in COMMON => 0

Read pages…………….

Successfull comparison! See onlyone,onlytwo and COMMON files

Files:

[treb2.txt]=231 words [rt2.txt]=181 words All=412

[onlyone]=223 [onlytwo]=173 [common]=8 All=412

Files metrics is correct

first Equal metric = 0.039415 [3%] –>Hihg Null-Equal metric = 0.038835 [3%] –>Medium second Equal metric = 0.019802 [1%] –>Down Medium = 0.032651 [3%]

Thus, when the requirements file is truncated to the keywords in the specialty, the result of the comparison becomes very convincing; when the dictionary of the first file (vacancy) is truncated to the keywords– terms that characterize the employee’s competencies, the methodology that we propose may be applicable for selecting job announcements with resumes that are suitable for the respective vacancies.

4. Summary

The described technique intended for the formation of a textual array’s semantic kernel makes it possible to compile a stable database of vacancies and a corresponding resume database. Automatic matching of texts will be effective if it is performed on the basis of an expertly determined basic semantic structure of the test being analyzed.

PART FOUR

METHOD OF QUALIFICATION COMPARATIVE ANALYSIS OF REQUIREMENTS TO PROFESSIONAL SKILLS

1. Main definitions

Nowadays there are practically no established feedback mechanisms that would allow in the shortest possible time to adapt educational programs to the rapidly changing demands of both the market and public sector of the economy. To solve are the problems of analyzing human capital and achieve a more rational use of human capital by improving the quality of training of qualified specialists and reducing educational costs by excluding outdated educational programs and correcting obsolete disciplines. It is therefor necessary to create and apply methods and tools for qualification comparative analysis of current labor market requirements and professional skills of specialists.

The presence of sufficiently large and existing databases of resumes and vacancies allows us to create techniques that are important both for choosing the future sphere of professional activity and for identifying current inconsistencies in the qualifications of specialists and requirements of the labor market. Based on the database of the website HeadHunter, the author has developed a method for determining the inconsistency of the skills of specialists and the requirements of employers to the specialty.

Qualification Comparative Analysis (QCA) is a procedure or process of identifying the difference between the current requirements of employers and the average level of current professional skills of applicants considered in a particular profession or specialty.

In course of QCA, on the basis of the set-theory comparison of dataarrays about the qualifications and skills of specialists and the requirements of employers, the area and measure of congruence, as well as the parameters of incongruence, are calculated.

Area of congruence is the area of intersection of arrays containing data on the professional skills of specialists and the requirements of employers, as a result of their set-theoretic comparison.

Measure of congruence is an integral category used in course of QCA for the numerical estimate which indicates the compliance of current applicants' job skills with the current requirements of employers.

Incongruence parameters determine the discrepancy between professional skills and the requirements of employers, expressed numerically or textually.

Objectives of qualification comparative analysis:

1) to identify problematic, from the point of view of specialists and employers of a particular field and specialty, aspects of work (unnecessary, outdated skills, unmet requirements regarding the "social package", etc.);

2) determine the range of required professional (including the so-called "soft skills") and personal skills and qualities that are not available to the specialists in this field and specialty;

3) to monitor the area and measure of congruence of professional qualification and requirements within the specialty.

2. Description of qualification comparative analysis technology

The method of identifying inconsistencies in the skills of a specialist, presented in the form of an unstructured resume text, to the requirements of the employer, drawn up in an arbitrary manner, is discussed in detail in article [22].

The purpose of this work is to show the method of identifying the measure of congruence and the incongruence parameters of the skills of specialists and the requirements of employers in the profession or specialty scale with the possibility of later making adjustments to national education programs.

To date, a comparative qualification analysis can be carried out using the complex of text indexing and analysis (CTIA), developed by prof. A.Yu. Shcherbakov and determining the degree of similarity of texts as a result of their set-theoretic comparison.

In the apparatus of non-selective mappings used for comparison of unstructured texts, the algorithm H (W) = h is used, which transforms a word of arbitrary length into a word of fixed length. Thus, the text becomes a sequence of hash words (binary numbers), each hash value | h |, therefore, any constructions for comparison have an equal length, which qualitatively changes the process of comparing texts in the direction of automation. A measure of the contents of the text indexed in this way by the m_ind program is Dictionary D. Then the indexed texts are compared by the tcmp program, which creates 3 files:

onlyone.csv - words found only in the first text

onlytwo.csv - words found only in the second text,

common.csv - words that occur in both the first and second texts.

Three measures of similarity of texts - maximum, natural and minimum are described in detail in article [21].

As objects of comparative qualification analysis, we take a set of employers' requirements in the form of vacancies (array A, treb.txt file) and a set of skills of specialists in the form of resumes (array B, rez.txt file). The indexing program m_ind creates the index dictionary, the treb.txt and rez.txt files, which are then compared by the tcmp program. As a result, we get three files:

common - a dictionary of words contained in both the requirements file and the summary file;

onlyone - a dictionary of words contained only in the treb.txt requirements file;

onlytwo is a dictionary of words contained only in the resume file rez.txt.

Along with the creation of these three dictionaries, the program gives a numerical assessment of the compliance of the skills of specialists with the requirements of the employer, which is their measure of congruence.
The QCA procedure is presented in general in the figure 1.

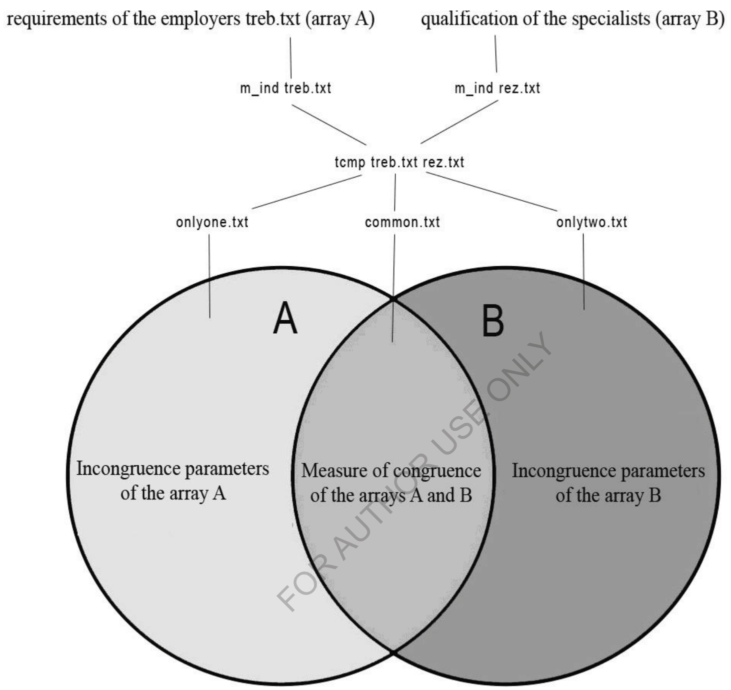


Figure 2 Application examples for qualification comparative analysis technique

3. Application example: QCA in the specialty programmer 1C

The application of the QCA procedure for comparing the array of actual requirements of employers to applicants and the corresponding array of resumes will be considered in such a widely-spread specialty as 1C Programmer. For comparison we have two files: trebpr ~ 1.txt, which contain employers' requirements (20 current vacancies), and rezpro ~ 1.txt, which contains resumes (20 current resumes) of specialists downloaded on hh.ru.

Comparing through the use of the complex of text indexing and analysis gave the following results:

Min word length in COMMON => 0

Read pages ....

Success comparing! See onlyone, onlytwo and COMMON files

Files:

[trebpr ~ 1.txt] = 816 words [rezpro ~ 1.txt] = 442 words All = 1258

[onlyone] = 681 [onlytwo] = 307 [common] = 135 All = 1258

Files metrics is correct

1-st Equal metric = 0.235436 [23%] -> Hihg

Null-Equal metric = 0.214626 [21%] -> Medium

2-d Equal metric = 0.120214 [12%] -> Down

Medium = 0.189902 [19%])

Table 1 shows coparison resultant files' fragments of trebpr ~ 1.txt and rezpro ~ 1.txt files (files common, onlyone and onlytwo)

Table 5. Requirements-skills coparison result fragments in the specialty Programmer 1C

Common	Onlyone	Onlytwo
Customization	Testing	Css
instructions	tasks	checks
business	technical	facial
optimization	jenkins	oracle
systems	uml	tax
revision	BP	Design
mistakes	tuning	reference
By the company	methodologies	descriptions
administration	Automation	Systemic
lead	ado	reporting
crm	layouts	Telecommunication

Revision	Escort	s
escort	services	administrator
Implementation	analyze	connection
bases	testing	operating
accounting	Study	Programmer
by staff	Bitrix	organizations
development	testing	vb
mechanisms	xml	servervisual
counseling	distributed	Conversion
tasks	detailing	sqlms
update	subsystems	clipper
reports	managerial	Platforms
Development	integration	program
correction	json	Accounting
additional	consul	excel
of documents	edi	oracle
data	run	installation
1C	construction	reinsurance
Control	technology	accounting 1C
technical	requirements	loading
Training	improvements	Update
ZUP	applied	setting
enterprises	printed	extensions
repository	system	Engineering
reports	configuration	OS
Writing	Conversions	accounts
bases	Bsp	sales
Programming	desk	data1C
exchange	modified	precious
configurations	architecture	statistical

Manufacturing software automation Search rls sql Support erp introduction	receptions incidents efficiency analysis xdto automated eepc Modernization	cameral Intertelecom shopping accessBase control certificate Systems Security

Starting point. The main functions of the 1C Programmer are the configuration and administration of software products, the maintenance of 1C configurations: the enterprise, the creation of new and modification (and updating) of existing documents, reports, and accounting, operational and management accounting processing. The 1C programmer develops methods and forms of accounting at the enterprise, trains employees to work with existing and newly created documents and reports [23].

The results of the comparison program software CTIA. We can distinguish the following area of congruence and the incongruence parameters.

The area of congruence of trebpr ~ 1.txt (requirements array) and rezpro ~ 1.txt (skills array). Both requirements and the skills contain terms related to the main functions of the 1C programmer described above, such as *tuning, optimization, refinement, administration, maintenance, accounting, implementation, accounting bases, development of mechanisms, updating, correction, training, configuration, automation, support, rls* (Row Level Security - a mechanism for restricting access to data) and sql (structured query language). In addition to the configuration of the PM (production management) as part of the 1C system, as well as the mechanism for generating DCS reports (data composition system), the common file (intersection of arrays of requirements and skills) includes *crm* (Customer Relationship Management), which indicates the objectively existing and the realized need for the participation of the 1C programmer in the creation and maintenance of a customer relationship management system of a company or organization.

Incongruence parameters for the array of requirements trebpr ~ 1.txt. Analysis of the array of words contained in the result file onlyone (words found only in the array of employers' requirements) showed an objective, but practically unrealized need for specialists to know how to develop software (**skill deficit parameter**) First of all, this is indicated by the presence of methodology in the list of words: *testing, code review, tdd* (test-driven development - development through testing), *bdd* (behavior-driven development - development "through behavior"). In addition, knowledge of standards for electronic data interchange between organizations (*EDI*), competent use of the standard subsystems library (*SSL*) in the development of application solutions on the 1C platform, as well as certain software systems for software integration (*Jenkins*), for database management (*postgresql*) are needed.

The words *distributed, consul* (a system for service discovery and a distributed repository of key value), *DIB* (distributed database) show that the employer expects the applicant to have skills in a distributed database, branches of which are separate bases of 1C:Enterprise.

Incongruence parameters for the array rezpro ~ 1.txt. The presence in the array of words related to the field of information technology in general, such as *processing, telecommunications, exel, installation,* or the basic terms of the specialty (*1C accounting, 1C workflow*) characterizes to a greater extent the presence of basic skills and knowledge that should have a priori a certain area of knowledge and which are not mentioned by employers. Let us call the indicator for the presence of such words **parameter of prior skills**.

In the rezpro ~ 1.txt array can be identified skills which are obsolete or unclaimed by the employer. These words are following: *CSS* (Cascading Style Sheets, a style sheet language used for describing the presentation of a document written in a markup language), *Oracle* (database management system), data, middleware, business applications), *vb* (Microsoft Visual Basic is a programming language, as well as an integrated software development environment), *clipper* (application programming system in a database environment that includes a compiler of programs written in the language near DBMS language dBase III +) and others.

Since the names Programmer 1C) were excluded from the resumes's texts, the presence of the words *system administrator, programmer 1C, software engineer, programmer-consultant* in this array indicates that applicants have experience in both analyzed and related specialties. In the array, there are practically no words denoting the names of positions that do not relate to the specialty 1C programmer (**job matching parameter**).

The measure of congruence of employers' requirements and skills of 1C programmers (the maximum measure of congruence is 23%) suggests that the specialty under consideration, despite its high relevance and apparent conservatism, is developing dynamically and requires deep knowledge of the employer's business context including principle outlines of the business process engineering.

Conclusions and recommendations. According to the results of a qualification comparative analysis of the compliance of current skills of 1C programmers with the requirements of employers, recommendations can be made for integrating software development methods, principles of building and operating of distributed systems and software systems for software integration, as well as for automation and notation techniques into educational programs in this specialty. In educational programs in this specialty, it is advisable to consider as obsolete vb (Microsoft Visual Basic is a programming language, as well as an integrated software development environment), clipper (application programming system in a database environment including a program compiler written in a language close to the DBMS dBase III +).

4. Application example: QCA in the specialty systems analyst

Let’s analyse the employers’ requirements in the specialty system analyst. On the website hh.ru 20 current vacancies and resumes containing employers 'requirements and applicants' skills (in unstructured form) were downloaded. Comparison of two arrays - the requirements of employers trebsy ~ 1.txt and the skills of specialists rezsys ~ 1.txt - gave the following results.

Min word length in COMMON => 0

Read pages ....

Success comparing! See onlyone, onlytwo and COMMON files

Files:

[trebsy ~ 1.txt] = 1044 words [rezsys ~ 1.txt] = 565 words All = 1609

[onlyone] = 834 [onlytwo] = 355 [common] = 210 All = 1609

Files metrics is correct

1-st Equal metric = 0.286415 [28%] -> Hihg

Null-Equal metric = 0.261032 [26%] -> Medium

2-d Equal metric = 0.150107 [15%] -> Down

Medium = 0.232286 [23%]

As they can see, the number of words in the trebsy ~ 1.txt and rezsys ~ 1.txt files is proportional to the number of words in the requirements and resumes files in the specialty 1C programmer trebpr ~ 1.txt and rezpro ~ 1.txt, respectively, therefore, you can compare their measures congruence.

Table 2 shows coparison resultant files’ fragments of trebsy ~ 1.txt and rezsys ~ 1.txt files - common, onlyone and onlytwo.

Table 6 Fragment of the comparison resulting files of the employer requirements and skills of specialists in the systems analyst specialty

Common	Onlyone	Onlytwo
Standards	Monitoring	Business
processing	Testing	optimization
methodologies	backlog	citrix
needs	mokap	radmin
testing	xsd	archiving
xml	uml	Escort
mistakes	api	visualparadigm
product	software	vs
integration	methodologies	Administration
Development	study	Technician
algorithms	to model	Consumer
atlassian	notation	calculations
jira	swagger	risks
development	visual	redmineatlassian
scripts	crm	gpo
systems	json	working out
notations	Escort	request
analysis	dwh	investment
writing	graphql	insurance
techniques	decomposition	dynamics
Update	modeling	change
bpmn	prototyping	cards
agile	esri	anaplan
sql	rabbitmq	support
analytics	gitlab	Leasing
control	being developed	factoring
confluence	support	performance

introduction	formats	platforms
skills	writing	analyst
custom	arcgis	configuring
implementation	seo	motorwert
maintaining	logic	
	service	
	olap	

Starting point. In general, the duties of a systems analyst include identification, collection, systematization, analysis of technical requirements included in specified sources, as well as their coordination and documentation in a specialized system, writing technical specifications for the development and modernization of the system [24]. The systems analyst creates the concept and architecture of the system, develops requirements for its individual functions, performs functional and logical design of systems. The responsibilities of the system analyst include the analysis, modeling and decomposition of business processes, the development of regulations for the implementation of activities, the construction of structural and functional models of the subject area of activity, the analysis of performance indicators to optimize them.

The results of the comparison with the use of the software CTIA. Here we will highlight the following area of congruence and incongruence parameters.

The area of congruence of trebsy ~ 1.txt requirements arrays and rezsys ~ 1.txt summaries. As it can be seen from the file containing the words and phrases that are in both the first and second texts, employers really need the applicants' skills to create and use software systems (this indicate the words *uml, task, mocap, sql, jira, testing, product errors, integration, development, writing of algorithms, automation, DBMS,* etc.), as well as ability to analyse and model business processes (*business project, project management, ability* *to write project documentation according to state standarts, TK for developing scenarios in notation systems, bpmn* - business processmodel and notation, *writing, techniques, automation, confluence, agile, backlog, design, writing techniques, decomposition, integration,* etc.)

Incongruence parameters of the trebsy ~ 1.txt (requirements array). However, as a result of comparing the two arrays of employers' requirements and the resume of the applicants, a significant discrepancy between the actual skills and the requirements was found. In particular, words and phrases like *Json, api, dwh* (Data Warehouse), *graphql, rabbitmq, gitlab, development, innovative applications, programming, silver light, seo, enterprise, angular, olap, figma, diagrams, idef0, prototyping , notation,* indicate the absence of the programming skills of applicants and the use of certain specifications and frameworks, as well as methods of functional modeling of business processes (idef0), storage management system and error tracking gitlab (**skill deficit parameter**)

Words and phrases editor*, contract monitoring, maintenance, logistics, design, swagger, gaap, Github, arcgis* indicate that the applicant will have to perform duties that are not directly related to the system analyst specialty.

Incongruence parameters of the resume array rezsys ~ 1.txt. Words and phrases that are found only in the second array (resumes of applicants), in turn, can also be divided into the following groups.

The content of the first group determines **absence of demand parameter**. First of all, these are obsolete skills. This group includes the words *citrix* (software solutions for virtualization, building computer networks, cloud computing), *radmin* (remote PC administration program), proxy (server intermediary between the user and the target server), *azureus* (cross-platform software for working with file-sharing networks via BitTorrent protocol), *dynamics*, *anaplan* (cloud platform for financial and operational planning and business process modeling), *soapui* (tool for testing web services), *redmine* (open server web application for managing projects and tasks), *powerdesigner* (a tool for creating business applications, including tools for business processes modeling).

There may also be words indirectly related to the requirements of the employer. For example, the visual paradigm tool from this array is used in agile projects that are present in both the first and second arrays. Let's call the indicator of the presence of such words as **the connectedness (binding) parameter.**

Given that specialists in their resumes, as a rule, indicate previous work experience, we can distinguish a group of words denoting a position, or a duty, or an area in which they worked previously. Such phrases as *system administrator, assistant systems analyst,* form a group of related specialties (job matching parameter), while the words *consumer credit, car loan, technician, manager, state databases, lessee, editing, scoring, credit rating , archiving, guarantee* explicitly indicate the presence in the resume of specialists of non-professional experience (**parameter of non-conformity of professional experience**).

It should be noted that information about experience in non related areas may be only an addition to other parameters, since the statistically chronological experience of applicants on average significantly exceeds the required experience for vacancies. However, on the basis of this information, conclusions can be drawn about the growing popularity of the specialty or the lack of applicants' qualification for a certain vacant position.

Conclusions and recommendations. With the development of information technology, the systems analyst's toolkit is also being improved, in order to take advantage of this a systems analyst should have knowledge in the methodology of functional modeling of business processes, which is confirmed by the analysis of the employers' requirements' array. It seems that along with the notation systems for describing business processes, the mastering of basic programming skills should be included in national education programs.

5. Summary

The presented methodology of qualification comparative analysis of the current requirements of the labor market for professional skills of specialists allows to determine their measure of congruence and incongruence parameters automatically and can be successfully applied to correct the national educational programs. The current and long-term comparative qualification analysis can provide not only significantly more rational use of human capital by improving the quality of training of qualified specialists, but also reduce the cost of education by eliminating outdated and outdated disciplines from educational programs.

The current qualification comparative analysis can be carried out by an automated analyzer - a pseudo-intellectual learning system or an artificial intelligence assistant associated with a search engine and processing data on employers' requirements and skills of specialists in real time. In the process of accumulation, processing and analysis of an automated analyzer of data on qualifications and requirements, the problem of determining the compliance of the subject area with the current state of the economy is solved. However, this problem requires further reflection and study. In general, the presented method successfully allows to generate new knowledge concerning the compliance of specialists' skills with the requirements of employers.

COMMON CONCLUSION

The first part of the work, devoted to the analysis of the existing experience of using the technology of distributed ledgers, its positive and negative effects and the practice of using it in various fields, allows us to determine the approaches to the first and second, as well as partly the fourth and fifth groups of problems of information processes in science, described in the conceptual introduction .

Recall that the first group of problems are technical problems associated with different linguistic environments and different knowledge levels of the languages used for scientific communication, in addition, not all researchers have equally convenient access to both scientific and information infrastructure. The second group of problems is the use of various software products by researchers, ranging from text editors with incompatible file storage formats to various email systems and messaging systems. The third group icludes problems of methodology, related both to the initially different approaches of researchers to the subject of research, and to different formats of the presentation of scientific material, for example, a different structure of articles and preprints, as well as different national systems for enumeration and registration of scientific results. The fourth group of problems is the reliability of the results and their correct time sequence, which is important for establishing scientific priorities and determining common points for the continuation and development of research in a given subject area. The fifth group of problems is related to the establishing “feedbacks” in scientific activity - updating the knowledge and skills of scientists and professionals involved in the technical support of scientific research, in accordance with the current level of both scientific knowledge, approaches and methods, and skills for their implementation. Remind that the fifth group connects the task of integrating scientific processes tightly with the education and training of spesialists of high qualification.

Thus, the main properties of the distributed registry technology considered in the second part of the paper for making informed decisions about the choice and use in order to ensure scientific and information activities solve the second and third group of problems, as well as partly the problems of the fourth and fifth groups.

The third part of the monograph, devoted to the technique of distinguishing the semantic core as a universal tool for the classification and systematization of unstructured information, allows us to solve the problems of the the third and fifth groups.

The fourth part of the work, devoted to the methodology of comparative qualification analysis, its application in order to achieve a more rational use of human capital by improving the quality of training of qualified specialists and reducing the cost of education problems also allows you to find better approaches to solving the third, fourth and fifth groups of problems.

The author, not claiming to fully cover the problems the scienceface in terms of the integration of information processes, nevertheless believes that the approaches proposed in the work are useful for the practical solution of the key problems that were listed above.

REFERENCES

1. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. – URL: bitcoin.org.

2. Schlatt, V., Schweizer, A., Urbach, N., and Fridgen, G. 2016. Blockchain: Grundlagen, Anwendungen und Potenziale. Projektgruppe Wirtschaftsinformatik des Fraunhofer-Instituts für Angewandte Informationstechnik FIT

3. Peters, G. W., Panayi, E. and Chapelle, A. Trends in crypto-currencies and blockchain technologies: A monetary theory and regulation perspective. 2015. – URL: http://arxiv.org/pdf/1508.04364.pdf.

4. Romano, D. and Schmid, G. Beyond Bitcoin: a critical look at blockchain-based systems. 2017. -URL: http://www.mdpi.com/2410-387X/1/2/15/htm

5. Ahamad, S., Nair, M. and Varghese, B. A survey on crypto currencies. Proceedings of the 4th International Conference on Advances in Computer Science (ACS 2013), December 13-14, Delhi, India.

6. Diffie, W. and Hellman, M. E. (1976) New Directions in Cryptography, IEEE Transactions on Information Theory, 22, 6, 644–654.

7. Mougayar William. The business blockchain. Promise, practice, and application of the next internet technology. 2016.

8. D.I. Pravikov, A.Yu. Shcherbakov. On the issue of changing the paradigm of information security // High Availability Systems – M: №2, т.14, 2018 – с. 35-39

9. Shermin Voshmgir. Blockchains, smart contracts und das dezentrale Web. 2016. Technologie Stiftung Berlin.

10. A. A. Ryazanova. The Blockchain Technology in Scientific and Information Activities // Scientific and Technical Information Processing, 2018, Vol. 45, No. 2, pp. 70–74.

11. Johannes Scherk B.Sc., Mag. Gerlinde Pöchhacker-Tröscher. Die Blockchain – Technologiefeld und wirtschaftliche Anwendungsbereiche. Pöchhacker Innovation Consulting GmbH. 2017.
12. Distributed Ledger Technology: beyond block chain. UK Government Office for Science. 2016.
13. Mainelli, M., von Gunten, C. Chain Of A Lifetime: How Blockchain Technology Might Transform Personal Insurance. 2014. A Long Finance report prepared by Z/Yen Group.
14. McKinsey & Company. Beyond the Hype: Blockchains in Capital Markets.
15. Biktimirov, M.R., Domashev, A.V., Cherkashin, P.A., and Shcherbakov, A.Yu., Blockchain technology: Universal structure and requirements, Autom. Doc. Math. Linguist., 2017, vol. 51, no. 6, pp. 235–238.
16. Digitale Wissenschaftskommunikation—Formate und ihre Nutzung, Gloning, G. and Fritz, G., Gießen: Gießener Elektronische Bibliothek, 2011.
17. Zaitsev, A.V., Gostev, S.S., Cherkashin, P.A., and Shcherbakov, A.Yu., Regarding the technology of distributed storage of confidential information in centers of general-purpose data processing, Autom. Doc. Math. Linguist., 2017, vol. 51, no. 3, pp. 117–119.
18. Recommendations for choosing a data processing cen-ter in Russia. https://habrahabr.ru/post/246419/.
19. Distributed Data Storage: From Cloud to Block. https://forklog.com/raspredelennoe-hranenie-dannyh-ot-oblaka-do-blokchejna/.
20. Where Do I Store Data for Decentralized Applications on a Cluster? https://habrahabr.ru/post/327836/.
21. Ryazanova, A.A. and Shcherbakov, A.Yu., To the problem of metrics of similarity of tests for methods of their automated comparison, Tekhnicheskie nauki: Nauchnye prioritety uchenykh. Sb. nauchn. tr. po itogam mezhdunarodnoi nauchno-prakticheskoi konferentsii (Technical Sciences: Scientific Priorities of Scientists. Proc. Int. Sci.-Pract. Conf.), Tolyatti, 2017.

22. A.E. Anisimova, A.A. Ryazanova, A.Yu. Shcherbakov. A Semantic Kernel as a Universal Instrument for Classification and Systematization of Unstructured Information in the Field of Human Capital // Scientific and Technical Information Processing, 2017, Vol. 44, No. 4, pp. 273–279.
23. Job description 1C programmer: duties, rights and responsibilities. – URL: https://businessman.ru/doljnostnaya-instruktsiya-programmista-s-obyazannosti-prava-i-otvetstvennost.html
24. Systems analyst: description of functions, position, specialty. – URL: http://hsbi.hse.ru/career/professions/sistemnyy-analitik/