AUTOMATION ALGORITHM OF THE INFORMATION SEARCH AND FIXATION IN UNFORMALIZED MESSAGES FOR THE MANAGEMENT OF SOCIALLY SIGNIFICANT INTELLIGENT PROJECTS

© 2022 D. O. Zhaxybayev[1]✉, M. N. Bakiyev[2]

[1]Zhangir khan West Kazakhstan Agrarian Technical University
51, Zhangir khan Street, 090009 Uralsk, Kazakhstan
[2]L. N. Gumilyov Eurasian National University
11, Pushkina Street, 010008 Nur-Sultan, Kazakhstan

Annotation. The article discusses the importance of automation for managing the intelligent processes and evaluating the effectiveness of processes. Often this process requires the creative approach of an analyst who has to read a lot of media information relating to the crowdsourcing project activity. This may be messages from users, persons concerned, experts` comments and other information generated by society as a response to the project activity. For many reasons, this is a non-optimal process, firstly, an analyst must review the entire array of information to estimate the project activity, and it is difficult to parallel this task. Since the social situation is subject to variation just for several hours and the human factor keeps away from such efficiency, then the evaluation relevance and up-to-dateness are compromised. It is necessary to automate the evaluation process of the crowdsourcing activity and this is the essence of our approach. We have developed the hierarchic thesaurus for highlighting the information from users` and persons` concerned messages, in the press reports and blog comments. Findings show that crowdfunding projects are hybrid in nature and include the elements of crowdsourcing, crowdsensing, crowdfunding, crowdworking, and crowdsourced recruitment. We explored the possibility of automating the efficiency evaluation of various crowdsourcing working processes and proposed design solutions for the development of such systems. The preliminary results indicate that leading issues include the lack of financial guarantees and the likelihood of information leakage to competitors. Hence, the first priority is to manage the exchange of money and information. By following this strategy, the crowdsourcing platforms will reach a high level of responsibility among project initiators and participants and will reduce the likelihood of tax avoidance by individuals, who received a financial reward.

Keywords: Crowdsourcing project, Crowdsourcing, Crowdfunding, Survey, Social network.

INTRODUCTION

The problem of the crowdsourcing project successfulness is far beyond the limits of managing or marketing challenges; this is enough comprehensive process which involves social, psychological, legal and other problems. For example, the rational choice hypothesis can be out in the case of information and communication technologies, clearly useful crowdtechnologied can be differently adopted by the end user and the adoption is not always associated with the user's usefulness [1]. To successfully implement a crowdsourcing project, we suggest using intelligent management automation based on a generated thesaurus, which creates a derivative problem — creating a thesaurus based on crowdsourcing. Thus, our work will solve a twofold problem: we will consider a thesaurus not for processing crowdsourcing data, and this thesaurus will be created based on the principles of crowdsourcing. One of the major problems in management is that the slow pace of technology adoption does not meet the

✉ Zhaxybayev Darkhan O.
   e-mail: darhan.03.92@mail.ru

burst of technological advancement and innovation. Attempts to make use of crowdsourcing as a means for increasing the efficiency are directly related to changes that occur in the information society and knowledge-based economy over time as well as to the advent of novel electronic communications. On the one hand, this is due to the variety of definitions of crowd-sourcing, and on the other - to the fact that most of these definitions use the language of management and concepts that do not have unambiguous correspondences in the thesaurus of the social sphere. Secondly, at the moment it is not clear how to systematize the use of crowd-sourcing in applied social research. Various examples of crowdsourcing classifications are given in the literature, but they also carry the problem of "foreign" language [2].

A well-built thesaurus has long been recognized as a valuable tool in the efficient operation of an information retrieval system. This paper presents the results of preliminary research aimed at determining the validity of the approach to the automatic construction of global thesauri (originally described by Crouch in [3] based on crowdsourced clustering of the collection of documents. We`ll show that the use of thesauruses created by this method leads to a significant increase in search efficiency in four test collections. The term "discrimination value theory" used in the thesaurus generation algorithm for determination of the membership the term in the certain thesaurus class disagrees the useful criterion for distinction between "good" and "indifferent" or "bad" thesaurus. In conclusion, we propose an alternative approach to crowdsourcing thesaurus building which significantly simplifies the work on creation of thesaurus viable classes. Experimental results show that the described here alternate approach in some cases gives thesauri comparable in search efficiency to thesauri obtained by means of the first method at a far less cost.

Advantages of using crowdsourcing largely depend upon the nature of the task. In general, by implementing crowdsourcing technologies, businesses can [4]:

• save their financial, human, informational, energy and time resources;

• bring to bear on the knowledge, skills, experience and interests of the crowd and individual participants;

• bring their outputs to the public;

• build up contacts with the target audience;

• encourage others to participate in managerial decision-making;

• complete projects in a shorter period of time.

In recent years, crowdsourcing has been developing rapidly around the world, such that small and medium-sized companies raise a significant part of their funds via crowd projects. Due to the huge growth of information technologies, such as Internet of things, cloud computing and so on, various institutions continuously generate tremendous data amounts for various applications. Under such conditions, it is necessary to have a management mechanism for operating the large-scale data with minimum costs [5].

While crowdsourcing offers new opportunities for the best analysis and processing the user data, business entities that look into contributing to such crowdsourcing should proceed with caution and stay informed to minimize the risk of data security violation [6].

The global crowdfunding industry grew almost 13-fold in four years, reaching 34.4 billion dollars in 2015 from 27.2 billion dollars in 2012. Hence, the annual increment is over 200 percent [7]. Approaches to managing data flows that are essentially very different have the different nature-different, in the works of many authors, the models of operating unstructured and not uniform data are developed "Creating a portable, high-level graph analytics paradigm for computer and data-intensive applications" [8].

In our work, we`ll develop a model for data extraction from documents suitable for management control of crowdsourcing processes in real time.

According to the 2016-2020 Global Crowdfunding Market Report, the crowdfunding market is projected to grow at the compound annual growth rate (CAGR) of 26.87 percent during the period 2016-2020 [9]. According to Technavio's analysts, it is expected to register a CAGR of almost 17 percent by 2021 [10]. Among the crowdfunding categories, the most popular include business and entrepreneurship (41.3 percent), social causes (19 percent), films and performing arts (12.2 percent) [7].

The first-category projects are funded via donations, in return for a reward, while social causes and films are financed under schemes where reward is not implied. In spite of its international nature, 70 percent of the global crowdsourcing industry is occupied by the US and European platforms. For instance, the top crowdfunding players as of 2018 are the North America (1.725.000.000 dollars) and Asia (1.054 billion dollars), followed by Europe (6.480.000.000 dollars) [9]. In that same year, P2P lending, which is financing tool that allows individuals to lend money to other individuals, became the most demanded form of crowdfunding, making up two thirds of the market (25 billion dollars). In 2016, the P2P lending segment accounted for almost 70 % of the global market and is projected to decline to 54 % by 2021, exhibiting almost 16 % decrease in market share [10].

Considering the above, the present research is relevant and of interest.

1. LITERATURE REVIEW

Conventional thesaurus in the information search environment can be called the global thesaurus. The term "global" is used here to distinguish this type of thesaurus from the main thesaurus [11]. In the global approach, thesaurus classes after creation are used for indexing both documents and enquiry messages. In contrast, the local thesaurus uses information obtained from documents received in response to a specific enquiry message for modification of this enquiry message which is then resent to the search system for processing [12]. Thus, a global thesaurus is created prior to the indexing process and is used to index both documents and queries, whereas a local thesaurus is created dynamically during query processing and is used to modify only that query [13]. This article explores various methods of constructing a thesaurus that have appeared in the literature, and describes an approach to the automatic construction of a global thesaurus based on the model of the values of discrimination of terms and the corresponding clustering algorithm [14]. The term "discrimination value model" itself is based on the vector space model, which will be described below.

The term "crowdsourcing" was first mentioned in [15] in 2006. As of now, a considerable amount of empirical material is accumulated, which allows for determining the main crowdsourcing trends. In [4], a general crowdsourcing model was presented. The said model is based on a holistic approach to the adaption of innovations in a company. One of the aspects embraced is the compatibility between different configuration choices of elements within a crowdsourcing project. Another work [16] examined the relationships between all components of the project: the crowd, the crowdsourcer, the task to be solved, and the crowdsourcing platform. In the meanwhile, there is a lack of engineering methods and principles that help create quality crowdsourcing-based solutions.

Today, researchers have not yet reached a consensus on exactly which practices fall within the category of crowdsourcing. In spite of disagreement, several of key characteristics have been identified [17]. Among them, crowdsourcing involves the distribution of tasks among a significant number of people (crowd). Although communication between participants traditionally occurs via the Internet, crowdsourcing does not necessarily mean the use of Internet services. Participants in crowdsourcing may receive monetary rewards, recognition, or skill development training. Sometimes, the results become aggregated and the choice is made in the favor of the best.

Depending on the context [18], crowdsourcing:
• provides for the transfer of certain production functions to an indefinite number of persons, and constitutes a public offer, which does not require an employment contract to be signed);
• is mainly used to collect and analyze aggregate data from a crowd;
• refers to the collaboration of people who voluntarily pool resources, mostly financial, usually via the Internet, to support other people or organizations in their efforts;
• helps companies to execute serious tasks by attracting people from all over the world, without making them leave their homes;
• provides means to select the most promising carriers of professional competencies, that

is, through observation and evaluation of candidates within the workflow-like context;

• relates to loan-based forms of donation, which involve individuals lending to other individuals or companies using specially created platform.

Exploring crowdsourcing projects is difficult in terms of methodology. In [19], authors draw our attention to that in the theoretical and practical aspects there is still "terminological chaos." In this connection, many aspects cannot be found in the theoretical framework. Many crowdsourcing definitions or interpretations do not contribute to the development of adequate techniques, which can be used to evaluate and compare different crowdsourcing projects. However, those methods that exist also need to be analyzed, and any limitations and inaccuracies that are inherent to them must be abolished.

Various forms of crowdsourcing were considered in [18], alongside a discussion on their use, which, as claimed, is aimed at turning the conventional content management applications into social machines in which tasks are performed as optimal combinations of human and computational intelligence. There is also a framework introduced for the analysis of the crowdsourcing systems. Factors that promote the development of crowdsourcing include the openness of companies towards ordinary users, the advantage of individuals' creativity in the innovation process, the creation of loyal virtual communities [20]. Kathleen Wilson and others were studying the crowdsourcing-based mechanism for increasing the efficiency of project management [21]. Considering the experience of five gamification companies in the US and Australia, crowdsourcing offers flexibility, speed, dynamism and scalability to project processes when the operation risks are low.

Taeihagh explored similarities and differences between different types of crowdsourcing and the sharing economy. The shared features that were established by him involve the use of information technologies, a reliance on crowds, monetary exchange, and the use of reputation systems [22].

Upon that, crowdsourcing campaigns are often launched for social causes [4], those in medicine [17], promoted by libraries, museums, archives [23], related to information products [24],

emergency management [25], and industrial projects [26].

In general, making decisions on crowdsourcing may involve risks, including legal risks that are associated with intellectual property. In this case [27], four different approaches may be applied to crowdsourcing intellectual property: passive, possessive, persuasive, and prudent.

As it is stated in [28], the crowdsourcing-based business model it tends to result in lower quality compared with traditional way of doing work, due to lose worker-employer relationships, skill diversity of the crowd and anonymity of participants. Also, it is crucial to verify characteristics of those involved in the crowdsourcing project, and to check the quality of their work at each stage [29, 30].

Vector space model. One of the main information search models is the vector space model. This model treats each document in the document collection as a set of unique words or word types. Then each document can be viewed as a vector of terms, and the complete collection of documents becomes a vector space of dimension m, where m is the number of word types in the collection. In the vector space model, the document vector, dj, is represented by a set of terms, where djk represents the frequency (or weight) of the term k in the document j (i. e. The number of times the term k appears in document j). If djk = 0, the term k does not appear in the dj document. Queries, as with documents, are represented by vectors of weighted terms.

Thus, despite the rapid pace of development, international crowdsourcing projects see a low level of responsibility on the part of project initiators and project participants. Additionally, a significant part of payments for the work remains off the book. The purpose of this article is to summarize the experience of participants in the crowdsourcing campaign, donors and initiators, to develop recommendations for the optimization of settings in which the crowdsourcing project takes place.

2. MATERIALS AND METHODS

Crowdsourcing was considered in our work as the combination of human and computation-

al intelligence, often involved when issues of managing the building software projects are actual. During this process, various tasks become available on the crowdsourcing platform, for example, knowledge interchange, problem solving, problem discovery and innovative solutions. This model is very useful when the team members are at work upon one problem because they know about the dynamics of developed solution, even persons with similar experience outside the group can be useful because processes related to individual experience, individual competence are often related to the problem. Under such conditions, the solution found by the crowdsourcing team core often simply pull a trump after improvement by a wider number of experts who were not involved in the direct development of the crowdsourcing product.

Before developing recommendations for the optimization of settings in which the crowdsourcing project takes place, it is advisable to investigate the experience of participants in such projects first.

With this aim in view, we conducted a survey of 447 respondents using computer-assisted telephone interviewing (CATI). The sampling process was performed in two stages. First, the overall population of Russia was clustered by regions, with the number of random respondents in the cluster group that is proportioned to the total number of residents in the given region. Residents were recruited through random digit dialing. Persons who admitted that they had experience of participating in the crowdsourcing project were asked if they want to proceed with online testing. Of those who agreed to participate in the survey, 211 (47.2 %) previously took part in the crowdsourcing projects. Of these, 183 (86.7 %) were acting as participants and donors, while the 28 respondents were project initiators.

Respondents were given a link to the online survey platform after receiving an informed consent to participate in the reward-based marketing research. In our opinion, such an approach is reasonable and allows collecting reliable data, since the proportion of Internet users is high enough to extrapolate research results to the entire population.

To validate data quality, the survey results were compared with the information gathered via profiling and in other projects. Multiple registrations were handled using the built-in platform features and limited to one for a certain time period. The originality of answers was also monitored. The sampling error did not exceed 5 %.

The survey results and recommendations developed on their basis are presented below.

3. RESULTS

An approach to building a thesaurus. When a user submits a request to an information system, the request is usually formulated to get as much relevant material as possible, while rejecting most of the irrelevant material. Thus, the user tries to formulate a query that provides both a high degree of recall (the proportion of relevant material extracted) and high accuracy (the proportion of relevant material extracted). Since recall and accuracy tend to vary inversely, such a result usually cannot be obtained. But information retrieval systems may try to improve the wording of the original query by providing methods to improve memories and / or improve accuracy. A thesaurus is generally considered to enhance recall because it expands the query by adding terms that are closely related or (in the context of a collection) synonymous with the original query terms. But a thesaurus can also be considered to increase accuracy if the classes of the thesaurus overlap or are non-exclusive [31]. Why is it so difficult to automatically build a global thesaurus? In addition to the current research efforts to create a relational thesaurus, almost all approaches to building dictionaries are based on the statistical coincidence of the word types in a text. The term "discrimination value model" tells us that only indifferent discriminators (which are generally low-frequency terms) should be grouped into thesaurus classes. However, term matching is particularly difficult to establish for low-frequency terms, where automatic thesaurus construction is most useful [32]. In fact, there are no reliable methods for using collection statistics to automatically build a thesaurus [33].

According to the 46th World Economic Forum, the key driver of the upcoming changes in

business will be the aligning with digital technologies. The crowdsourcing technology here serves a tool to accomplish this aligning and can be regarded as a contributor to the innovation-driven development [34]. Crowdsourcing technology is based on close social interaction of a large number of people at different levels. Development of an automated control and decision system support in this area will require the creation of an approach for automatic analysis of human communications. To develop the algorithm, it is necessary to use corpus systems, with the help of which you can build the cognitive structure of a certain semantic field (as an example, as often used in volunteer projects — environmental). Such structures will become the basis for the selection of keywords and phrases. Thus, we present the basis for graphical clustering algorithms. The next step is to consider the interrelationships of lemmas combined into lexical-semantic groups, synonymic series, antonymic pairs, and associative chains. This work contributes to the development of an algorithm for identifying related components. To obtain such structures, we select corpuses of small size, sufficient for applied purposes. The next step should be building the concordance. The statistical calculation program will allow to identify key vocabulary, including collocations (combinations of words), which is subjected to frame processing. Semantic grouping by conceptual features relies on the subjective experience of the author and the nominative picture of the considered fragment of reality available in the language. Thus, the compiled concordance will allow to perform hierarchical clustering. Analysis of the existing dictionaries in the environmental field shows their insufficient functionality, since they perform the function of a reference book. At present, the role of electronic dictionaries of this type is essential, when the user finds the necessary material at the request of his thought process to express some idea within the framework of the thematic field. Such a query determines the importance of a functionally meaningful dictionary based on a field cognitive structure with a hierarchy of primary and secondary links. When developing such a structure, we rely on our own experience, research the encyclopedic material, and identify aspects that will be essential for the

user (in this case, the user can be any member of the society or an expert in this field). In any case, such a dictionary, intended for the general user, will have a thesaurus that is known to everyone. The need to create a functional-thematic dictionary according to the new concept is due to the importance of understanding the account of the paradigmatic and syntagmatic connections of the nominative scheme (semantic network) of one field. Such a dictionary should have such features as logical coherence, structure, and informativeness. The first stage of the formation of this type of dictionary is the identification of a classification system with a hierarchically represented set of semantic blocks.

1. The first group is called "Living organisms". It includes concepts that have their own life cycle. In this categorical group, there are three main types: living creatures, plants, and microorganisms. The classification was based on sections from ready-made models that classified living organisms as a separate class. The nominative picture of the considered thematic field will serve as a support for the systematic representation of the thematic vocabulary. The links in this structure are lemmas that should serve as signals for information retrieval of the computer version of the dictionary. Thus, the semantic network of the basis of the future electronic dictionary distinguishes it from existing dictionaries in the field of ecology. The cognitive structure of the thematic dictionary will be built in such a way that the user can logically understand the relationship of all living beings on the basis of categorical, typical, generic, and species levels. At the same time, at each level, in the second position, the potential lexical-semantic compatibility (syntagmatics) will be revealed, indicating parallel variations (paradigmatics).

2. The second group is "Inanimate organisms". This categorical group has 6 subordinate groups, of which some are standard categories for ecology, and some were identified during the selection of terms by group. By the beginning of the classification, 4 standard categories were identified which, like the categories of the first group, were taken from existing categories. Standard ones include the following: 1) "Land" which includes

the terms "earth", "dry land", "soil", etc.; 2) "Water space" specified by names of all types of water bodies; 3) "Air", this category, apart from names of "air space", contains the words designating gaseous substances; 4) "Cosmic space" includes the names of all celestial bodies and the concepts of space. The analysis of all ecological concepts also shows that there are the names designating the living body particles, but they cannot be designated as living bodies themselves. They are divided into natural ones (for example, vitamins, hormones, homotelergones (the latter include biolins, epagones, odmichnions, toribones, gonofions, etofions)) and artificial derivatives as the result of human activity (we`ll refer drugs to them). Inanimate organisms also include objects that are products of human activity (premises, dishes, household items, etc.). In the planned dictionary, it is advisable to distinguish a group of lemmas-concepts that denote geographical objects such as the area, landscape, etc. A separate group should be placed in the semantic group "direction of ecology", in which each direction is served by a specific list of concepts.

3. The third group is "Being". The nominations described in the above groups do not exist in isolation, vital activity presupposes certain interrelationships both within one category and between objects and subjects of different categorical groups. Therefore, for the disclosure of the thematic field "Ecology", it is important to consider the universal concepts of being, relations, connections, states, characteristics, and actions. These groups are identified based on the analysis of the dictionaries of the environmentally oriented thesaurus. The most numerous class was "process", the others are approximately the same in terms of the number of terms "ability", "system", "properties", "state", "condition".

The best example is the category "process", because it is the largest in terms of the number of concepts. It is also easy to identify it by its main features. The first most understandable selection is a specific indication of the process in the text of the interpretation of the term. For example, the term "adaptation" has the following definition: "adaptation is the process of adaptation of organisms and species that occurs (during biocenogenesis)". The following criteria for identifying terms

in the "process" category are keywords, such as: assimilation, accumulation, change, movement, influence, occurrence, absorption, etc. For example, adsorption is absorption of various substances from solutions. Salinity of the soil — the accumulation of water-soluble salts that cause salt marshes and salinity of the soil. The remaining definitions were selected by understanding the text of the term, and if a process was described, the term was assigned to the appropriate category. At this stage of working on this category, we are considering the allocation of a subcategory, for a more specific definition of terms.

Using the system of building a dictionary on the principle of hierarchy, rather than the standard alphabetic system of representation of terms, we propose algorithms for the compatibility of thematic vocabulary based on the thesaurus.

Hierarchical and associative relationships can be described using a multi-level system of relationships. We will denote each connection by certain notations. The main symbols are TT, BT, NT, RT. The first designation TT (top term) is a term that is located at the top level of the hierarchy, in Figure 1 it is "living organisms". A term that has a broader concept, and stands one step higher, is denoted by BT (broader term), and if it stands one level lower, it is denoted by NT (narrower term). Associative relation uses the designation RT (related term).

The following offers how the terms interact between themselves.

For description of hierarchic relations shown in Figure 1, which was built for viewing the concepts interaction between themselves. The specific concept "bird" is presented by categorical concept "living bodies" because it is the highest category in this scheme. However, the same term "bird" has the closest specific concept given as "animal world", at the same time, it is lower-level specific concept for this category. While the syntagmatic relationship for this term "bird" can be the word "to fly", but this is only an example of one concept, in real dictionaries there may be a large quantity of such propositions or may not be at all. When considering this issue, we were based on the opinions of scientists who considered the concept categorization.
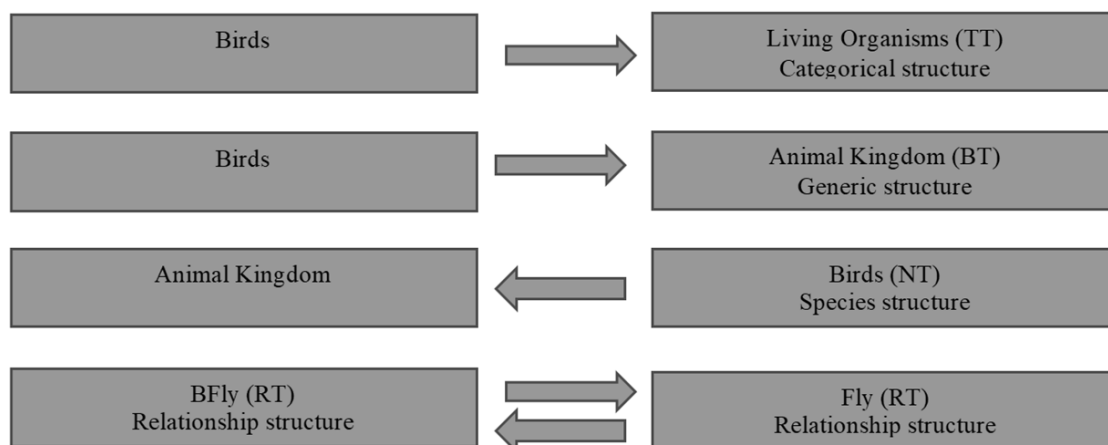
Fig. 1. Multilevel communication system

4. DISCUSSION

The crowdsourcing event is an ambiguous phenomenon. It can take place as part of economic (i.e., via investing activity as well as in business and financial management), socio-psychological (i.e., as group and consumer behavior of participants) and marketing (i.e., for product promotion, collection and analysis of feedback, advertising, and PR campaign) processes.

As web-technologies advance, many people take part in information production and distribution in the web-environment. In addition, many researchers concerned themselves with studies on elaboration of useful information using a thematic recommendation system, as the volume and complexity of web information run up [35]. The collective creation processes are not an exception, crowdsourcing as a collective creation practical activity existed earlier, but only in the recent decades crowdsourcing mainly bases itself upon network communications, resulting in a lot of opportunities and drastically expansion of the circle involved in creative activity. Thanks to spreading out the intelligent devices and growing the network technologies, everyone can easily get access to web-data. In addition, everyone takes an active part in information production and distribution and this suggests the rapid growth of the volume and complicity of web-data. Big data are classified as structured or unstructured data that cannot be collected or analyzed using conventional methods, the traffic within the crowd-sourcing community in operation belongs to the same data [35].

At the current stage of economic development, crowdsourcing is considered the leverage of a large group of people to do the work that was previously conducted by one specialist alone. The concept in question refers to the compliance with the open source principles in areas that not related to programming. The crowdsourcing campaign is viewed as a means of problem solving, which turns to numerous organizations for assistance, most often via the Internet. It is a business model within the framework of which a company outsources tasks to a third party or a group of consumers (i.e., a crowd). Making use of the Internet allows organizing the interaction between consumers in a simple way and without unnecessary expenses as well as approach tasks that were unimaginable in the era before the Internet. Crowdsourcing can strengthen the provider-consumer relationships. Using the advantages of a crowdsourcing platform, consumers can receive the desired services, ideas and any other content from the crowd, usually the online communities, rather than regular employees or suppliers. People who are acting independently are ready to provide their services in return for experience, recognition or reward. Modern-day employers understand that technological advances allow people to develop their skills and talents without leaving home, so they use social media and online forums where people are invited to participate in the projects. However, there are various security challenges and problems associated with remote work of employee team, working on a common task, if the work supposes the big data transfer [36].

Crowdsourcing is part of what an American economist and innovation specialist Eric Von Hippel calls the *user-centered innovation*. In his own words, "*Innovation is rapidly becoming democratized. Users, aided by improvements in computer and communications technology, increasingly can develop their own new products and services. These innovating users–both individuals and firms–often freely share their innovations with others, creating user-innovation communities and a rich intellectual commons*" [37].

Thus, crowdsourcing has undergone changes. Now, it is a modern management tool, which allows solving business problems by utilizing the knowledge potential of employees, customers, and partners.

The crowdsourcing platforms are to a greater extent one of the effective communication tools implemented within the framework of an innovative project, rather than a direct source of funding. If properly conducted, the crowdfunding campaign draws attention to the idea thereby increasing the chances of an innovative project to win the attention of consumers and investors. With this approach to crowdsourcing, it is the attracted audience feedback that is the indicator of success, rather than funds raised.

The potential of socially active collective mind is that it provides for the encouragement and engagement of Internet users in order to solve a clearly stated problem. In doing so, the collective mind finds new opportunities for building up multidimensional contacts with the society.

The concept of crowdsourcing has been criticized though, due to application- and management-related cons. The number one claim of crowdsourcing opponents is that the crowd is anything but smart; it is not able to create valuable things, as it consists of different people who will fail to agree upon and develop an effective solution. Most often, a solution rises when the crowd itself consistently rejects incorrect decisions and thereby settles on a positive result. The second disadvantage is the unfair reward. Only a limited number of people receive a reward, while the remaining participants work for free. This gives rise to a belief that professionals make little account of crowdsourcing technologies, since they value their own time and do not want to waste it.

However, almost all platforms for creative crowdsourcing (approach in which participants create complex labor-intensive products) suggest working on the basis of prototypes or limited solutions. Further work in this case is possible only when the customer defines one or several winners. For many professionals in developing countries, taking part in crowdsourcing projects may be a major source of income.

Third, the crowdsourcing solution is not necessarily the cheapest way of solving a problem. For example, hiring a logo designer to create a trademark is often a money-saving decision because resources allocation for logo showcase, communication with the audience, and advertising need additional funds to be put into action. The last two disadvantages are the information leakage and inconsistency in the actions of participants.

According to our observations, the major problems with the crowdsourcing campaigns are that the initiated projects are not elaborated to a sufficient extent and lack the levers of influence on participants in the project (to make them deliver on obligations) and on the initiators (to make them pay for the work). The increase in the information volume controlled by an information manager over the past decades influenced on the need to automate the processing of unstructured information in order to develop optimal management decisions. And last but not least, this is my contribution to increasing the confidentiality of the management process by using a hierarchical thesaurus approach to the processing of some confidential pieces of information, and excluding an analyst employee from this process. This improvement in the security of the project can be considered very important. An extensive overview of new cyberspace security challenges and processes in cyberspace is given in the book [38]. to ensure the security of processes in cyberspace, different authors offer very different methods, as an example, we can mention the work [39] suggesting the use of even deep learning neural networks to ensure the security of cyber processes; neural networks are a widely used tool for many computational tasks, but it seems for the first time that we see the use of this tool for ensuring transaction security, which can serve as an indi-

rect marker of the complexity of the discussed problems and an unsatisfactory assessment of the situation with the security of network protocols, since such exotic tools appear. As a similar exotic tool for analyzing the problems of interaction between a large number of individuals, the work can serve [40]. We consider this a promising approach that can be applied to develop the problems described in our work; metaheuristic algorithms are an example of adequate solutions to difficult-to-define problems, such as the problem of ensuring the safety of a team of employees involved in a crowdsourcing project.

To increase the crowdsourced security and the likelihood of crowdsourcing project implementation, platforms must deanonymize both the participants and the initiators of the crowdsourcing project. Consequently, the social networks will play a key role in the development of crowdsourcing, which was the reason for its emergence in the first place. In social communities, individuals exchange information on a constant basis and in doing so they create an opportunity for generating, distributing and discussing benevolent content. Through the rapid dissemination of information, social networks have the ability to mobilize human resources within a short timeframe to solve any social problem as well as provide the history of activity of individual participants and initiators.

That way, various websites and social networks, which are used to popularize the crowdsourcing projects and attract investors, resemble a unique phenomenon, as they transform the accumulated social capital into financial capital. The crowdsourcing platforms can also be used to implement research projects and undertakings. A crowdsourcing project is characterized by a broad range of subjects, microfinancing, cross-border engagement of participants, networked organization, and by cost-effectiveness [18].

5. CONCLUSIONS

We developed and conducted an experiment to study the possibility of building a global thesaurus based on a crowdsourcing model for determining the meaning of discrimination of terms and the corresponding clustering algorithm.

A thesaurus construction procedure was developed, implemented, and applied to two standard test collections. The results of the experiment show that the basic structure of the thesaurus classes can be disclosed for a particular collection by applying the appropriate clustering algorithm to the documents and recognizing certain relationships between the low-frequency terms of the clustered documents. Moreover, such a thesaurus can be built in the course of a crowdsourcing process; in the future, such a process can even be automated. The current crowdfunding projects are usually hybrid and include the current crowdfunding projects are usually hybrid and include the elements of crowdsourcing, crowdsensing, crowdworking, and crowdsourced recruitment

Some conclusions that can be drawn from the preliminary analysis using our proposed approach to the crowdsourcing project of environmental monitoring of salt-dome landscapes of East Kazakhstan, allows us to draw the following conclusions.

Due to the lack of contractual relationship between participants in the project and its initiators, the major problem that arises in crowdsourcing is the lack of financial guarantees. Additionally, participants may be involved in the leaking of project-related information to competitors at any stage of project realization, which places a question mark over the reasonability of crowd involvement. These problems can emerge alongside the secondary issues:

• valuable information leakage due to the actions of malicious executants;

• project abandonment;

• the lack of discipline among participants, especially after receiving an advance payment;

• the lack of mechanisms for influencing project participants;

• self-indulgent and unpredictable behavior of participants.

The act of deanonymizing participants and initiators by referring to their social pages will provide parties that are involved in the crowdsourcing project with the following information about the workforce standing behind the project:

1. Competencies and skills, from professional to specific such as:

• teamwork competence;

• soft skills;

• creativity

• systematic thinking, etc.;

2. Details about the personality:

• tolerance to constructive criticism;

• character traits (attentiveness, perseverance, diligence, being trustworthy);

3. Project participation/realization statistics.

From this perspective, managing money and information flows is a first priority to ensure the overall, financial and information security within the project.

To successfully promote crowdsourcing, initiators must to increase their communication with the crowd. For this purpose, mediators such as crowdsourcing agencies and platforms that combine functions of an information website and a financial intermediary may be involved. The most famous crowdsourcing platforms are the Amazon Mechanical Turk (MTurk) and the Inno Centive. There are also cases of capital rise via highly specialized intermediaries (e.g., using blog posts). In this case, initiator turns to the crowd with a specific problem (e.g., posts a message will all the insights needed), sets a timeframe and a certain reward for its solution. The initiator has no right to directly influence the crowdsourcing process.

The crowdfunding process differs from crowdsourcing in terms of mediators. The crowdfunding platforms provide for the accumulation of financial resources. The Kickstarter, for example, is a platform where project initiators can set the funding goal, the timeframe, and a reward for each specific project.

In crowdsourcing, participants in communication are engaged in the equality-based work relationship; hence, exploitation and propaganda are not an issue here, especially if there is an opportunity to make public the cases of non-payments, fraud, failure to meet deadlines, etc. Thus, removing confidentiality of information about participants in the crowdsourcing project and its initiators by referring their social webpages ensures both the information security and financial discipline. This approach is aimed at protecting the participants. Yet, services that help employees influence the level of wages and work arrangements are increasingly emerging.

In the context of virtual communication and weak formal institutions, the role of social ties is gaining importance, such that these ties become the main leverage to increase work efficiency and protect against precarization. The long-term cooperation relationships and contacts that were built up through the social networks result in trust between parties and guarantee the fulfillment of obligations.

Social networks are not subject to crisis and external factors, they do not depend on political, economic, environmental, or demographic situation in a country or region. The only threat here is a decrease in attendance; however, there is always an opportunity change the platform.

Social networks provide a detailed statistic about the use of links, articles, and content distribution. Moreover, they can be used to popularize the crowdsourcing projects and attract investors.

The crowdfunding platforms can set a pledge threshold so that the funds are transferred only upon meeting it but not every platform has such a feature. The use of a pledge crowdfunding mechanism depends on the choice of a business model.

We suggest considering the following indicators of the crowdsourcing campaign:

1. Goal Fulfillment Rate. Reflects the quality of project promotion;

2. Average Donation.

3. The Number of Donors.

4. The Number of Page Views.

5. The Number of Clicks that the Project Received. Reflects the success of a PR campaign;

6. The Proportion of New Investors. Reflects the success of project promotion;

7. Investor-to-Total Viewers Ratio. Allows the project initiator to evaluate the return on the innovation idea;

8. The Number of Comments Left. Projects may be of different interest to platform users.

9. Statistics collected for other projects that were launched by the same initiator(s);

10. Project Openness (this refers to the provision of information about the fundraising progress and professional views).

Project initiators can understand the opinion of potential buyers about the project idea by analyzing the tone in which they write. It is also crucial to respond to almost every comment, which they leave.

Crowdsourcing solutions are gaining popularity among small investors who seek a platform for the accumulation of small capital in return for a certain reward. Such investors will not become business angels, as they do not have sufficient resources for independent or share financing of innovations. Beyond that, the popularity of this method is pumped by a decrease in the public confidence in banks, which was triggered after some of them reached bankruptcy.

Today, researchers have not yet reached a consensus on exactly which practices fall within the category of crowdsourcing. According to different sources, the act of crowdsourcing involves the distribution of tasks among a significant number of people. Although communication between participants traditionally occurs via the Internet, crowdsourcing does not necessarily mean the use of Internet services. Participants in crowdsourcing may receive monetary rewards, recognition, or skill development training. Sometimes, the results become aggregated and the choice is made in the favor of the best.

Due to the lack of contractual relationship between participants in the project and its initiators, the major problem that arises in crowdsourcing is the lack of financial guarantees. Additionally, participants may be involved in the leaking of project-related information to competitors at any stage of project realization, which places a question mark over the reasonability of crowd involvement. These problems can emerge alongside the secondary issues such as the leaking of valuable information due to the actions of malicious executants; project abandonment; the lack of financial discipline; the lack of mechanisms for influencing project participants; self-indulgent and unpredictable behavior of participants.

Respondents in this study recommend companies to carefully weigh their decisions regarding the projects and to launch only well-elaborated projects. Aside from that, the crowdsourcing platforms must deanonymize the participants, especially in projects that offer financial rewards. Such an approach allows the increase in responsibility for the quality of work and financial matters.

Within the framework of our approach, based on hierarchical thesauri, models can be created for any fragment of reality (for all spheres of so-cial activity). Equally, algorithms can be created to automatically assess the attitude of a large number of individuals expressed in writing to any socially significant issue.

It remains to be determined whether these results can be generalized.

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Abdulfattah F. H.* (2019) Factors Affecting Students' Intention Toward Mobile Cloud Computing: Mobile Cloud Computing. International *Journal of Cloud Applications and Computing (IJCAC)*. 9(2). P. 28–42.

2. *Schenk E. and Guittard C.* (2009) Crowdsourcing: What can be Outsourced to the Crowd, and Why. *Workshop on open source innovation, Strasbourg*, France. 72.

3. *Crouch C. J.* (1990) An approach to the automatic construction of global thesauri. *Information Processing & Management*. 26(5). P. 629–640.

4. *Fuger S., Schimpf R., Füller J. and Hutter K.* (2017) User roles and team structures in a crowdsourcing community for international development – a social network perspective. *Information Technology for Development*. 23(3). P. 438–462.

5. *Sambrekar K. and Rajpurohit V. S.* (2019) Fast and Efficient Multiview Access Control Mechanism for Cloud Based Agriculture Storage Management System. *International Journal of Cloud Applications and Computing (IJCAC)*. 9(1). P. 33–49.

6. *Wolfson S. M. and Lease M.* (2011) Look before you leap: Legal pitfalls of crowdsourcing. *Proceedings of the American Society for Information Science and Technology*. 48(1). P. 1–10.

7. CF The Crowdfunding Industry Report (2015) Available at: http://reports.crowdsourcing.org/index.php?route=product/product&product_id=54

8. *Searles R., Herbein S., Johnston T., Taufer M. and Chandrasekaran S.* (2019) Creating a portable, high-level graph analytics paradigm for com-

pute and data-intensive applications. *International Journal of High Performance Computing and Networking*. 13(1). P. 105–118.

9. Global Crowdfunding Market 2016–2020, 2019. Available at: https://www.marketresearch-store.com/report/global-crowdfunding-market-2016-2020-45637

10. Global Crowdfunding Market 2017–2021, 2019. Available at: https://finance.yahoo.com/news/global-crowdfunding-market-2017-2021-182400851.html

11. *Azad H. K. and Deepak A.* (2019) Query expansion techniques for information retrieval: a survey. *Information Processing & Management*. 56(5). P. 1698–1735.

12. *Efthimiadis E. N.* (1996) Query Expansion. *Annual review of information science and technology (ARIST)*. 31. P. 121–87.

13. *Schuster R., Schuster T., Meri Y. and Shmatikov V.* (2020) Humpty Dumpty: Controlling word meanings via corpus poisoning. *2020 IEEE Symposium on Security and Privacy (SP). IEEE*.

14. *Campos R., Mangaravite V., Pasquali A., Jorge A., Nunes C. and Jatowt A.* (2020) YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*. 509. P. 257–289.

15. *Howe J.* (2006) The rise of crowdsourcing. *Wired magazine*. 14(6). P. 1–4.

16. *Hosseini M., Phalp K., Taylor J. and Ali R.* (2015) On the Configuration of Crowdsourcing Projects. *International Journal of Information System Modeling and Design*. 6(3). P. 27–45.

17. *Wazny K.* (2017) "Crowdsourcing" ten years in: A review. *Journal of global health*. 7(2).

18. *Simperl E.* (2015) How to use crowdsourcing effectively: Guidelines and examples. *Liber Quarterly*. 25(1). P. 18–39.

19. *Lenart-Gansiniec R.* (2018) Methodological challenges of research on crowdsourcing. *Journal of Entrepreneurship, Management and Innovation*. 14(4). P. 107–126.

20. *Zhao Z.* (2019) Idea Crowdsourcing for Innovation: Fundamentals and Recommendations. *Management & Data Science*. 3(2).

21. *Wilson K. B., Bhakoo V. and Samson D.* (2018) Crowdsourcing: A contemporary form of project management with linkages to open innovation and novel operations. *International Journal of Operations & Production Management*. 38(6). P. 1467–1494.

22. *Taeihagh A.* (2017) Crowdsourcing, sharing economies and development. *Journal of Developing Societies*. 33(2). P. 191–222.

23. *McKinley D.* (2012) Practical management strategies for crowdsourcing in libraries, archives and museums. Available at: http://www.digital-glam.org/crowdsourcing/crowdsourcing-strategies/

24. *Niu X. J., Qin S. F., Vines J., Wong R. and Lu H.* (2019) Key crowdsourcing technologies for product design and development. *International Journal of Automation and Computing*. 16(1). P. 1–15.

25. *Harrison S. E. and Johnson P. A.* (2016) Crowdsourcing the disaster management cycle. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*. 8(4). P. 17–40.

26. *Saldanha F. P., Cohendet P. and Pozzebon M.* (2014) Challenging the stage-gate model in crowdsourcing: The case of Fiat Mio in Brazil. *Technology Innovation Management Review*. 4(9). P. 28–35.

27. *de Beer J., McCarthy I. P., Soliman A. and Treen E.* (2017) Click here to agree: Managing intellectual property when crowdsourcing solutions. *Business Horizons*. 60(2). P. 207–217.

28. *Iren D. and Bilgen S.* (2014) Cost of quality in crowdsourcing. *Human Computation*. 1(2). P. 283–314.

29. *Naroditskiy V., Rahwan I., Cebrian M. and Jennings N. R.* (2012) Verification in referral-based crowdsourcing. *PloS one*. 7(10). e45924.

30. *Santos C. P., Khan V. J. and Markopoulos P.* (2014) On utilizing player models to predict behavior in crowdsourcing tasks. *International Conference on Social Informatics*. Springer, Cham. P. 448–451.

31. *Grossman D. A. and Frieder O.* (2012) Information retrieval: Algorithms and heuristics. *Springer Science & Business Media*. 15.

32. *Curran J. R.* (2004) From distributional to semantic similarity.

33. *Saidi F., Trabelsi Z., Salah K. and Ghezala H. B.* (2017) Approaches to analyze cyber terrorist communities: Survey and challenges. *Computers & Security*. 66. P. 66–80.

34. *Baeva O. N. and Malyshenko G. V.* (2017) On issue of essence and scales of using crowd-sourcing. *Baikal Research Journal.* 8(2).

35. *Kim M., Gupta B. B. and Rho S.* (2018) Crowdsourcing based scientific issue tracking with topic analysis. *Applied Soft Computing.* 66. P. 506–511.

36. *Gou Z., Yamaguchi S. and Gupta B. B.* (2017) Analysis of various security issues and challenges in cloud computing environment: a survey. *Identity Theft: Breakthroughs in Research and Practice.* IGI Global. P. 221–247.

37. *Von Hippel E.* (2009) Democratizing innovation: the evolving phenomenon of user innovation. *International Journal of Innovation Science.* 1(1). P. 29–40.

38. *Gupta B. B.* (Ed.) (2018) Computer and cyber security: principles, algorithm, applications, and perspectives. CRC Press.

39. *Jiang F., Fu Y., Gupta B. B., Lou F., Rho S., Meng F. and Tian Z.* (2018) Deep learning based multi-channel intelligent attack detection for data security. *IEEE transactions on Sustainable Computing.*

40. *Ramos J., Nedjah N., de Macedo Mourelle L. and Gupta B. B.* (2018) Visual data mining for crowd anomaly detection using artificial bacteria colony. *Multimedia Tools and Applications.* 77(14). P. 17755–17777.

Zhaxybayev Darkhan O. — Master of Pedagogical Sciences, Lecturer of the Department of Information Systems at West Kazakhstan Agrarian and Technical University named after Zhangir Khan.
E-mail: darhan.03.92@mail.ru
ORCID iD: https://orcid.org/0000-0001-6355-5431

Bakiyev Murat N. — Candidate of Physical and Mathematical Sciences, Acting Associate Professor at the Department of Information Systems at L. N. Gumilev Eurasian National University.
E-mail: murat26261957@mail.ru
ORCID iD: https://orcid.org/0000-0002-5131-4275