Integration of Social Media Platforms and Specialized Web Resources for the Effective Use of High-tech Medical Information

Alyona Kozyreva, Uliana Nazarenko, Grigory Shovkoplias, Artem Beresnev, Elizaveta Klevtsova and Natalia Gusarova
ITMO University, 49 Kronverksky av., St. Petersburg, Russia

Keywords: Chatbot, Healthcare, High-tech Medical Information, Internet Community Structure, Publicly Accessible Resource, Second Opinion, Spontaneous Crowdsourcing, Telegram Channels, VK Community.

Abstract: The widespread dissemination of information technologies and technological breakthroughs in recent years have led to the fact that the flow of medical information falls on the patient. In order to interpret their high-tech medical information, given the lack of doctors and mistrust in them, which is especially typical for developing countries, people make attempts to use available Internet sources and collective network intelligence, i.e. appeal to collective opinion. Thus, there is the problem of integration of social media platforms and specialized web resources for the effective use of high-tech medical information. We consider this problem in relation to the Russian-speaking segment of the Internet. We have experimentally studied the structure of public medical Internet communities typical for Russia. We found that they are characterized by self-organization. We have developed and launched a web resource for the effective use of high-tech medical information, and to form the motivational component of the resource, we use the identified structure of already existing network communities of medical focus. We use specialized chat bots as an effective means of integrating the developed resource and network communities.

1 INTRODUCTION

The prevention and treatment of top 10 diseases causing the most deaths worldwide, including coronary artery disease, stroke, chronic obstructive pulmonary disease, bronchus and lung cancers etc. (Pietrangelo, 2019) increasingly relies on high medical technologies. Incorrect or inadequate (without taking into account the health status in general) interpretation of such information by the patient himself can cause him psychological and social problems, not to mention the treatment of the disease itself (Giardina, 2018). However, the demand for such information from patients all over the world is only increasing.

There are several reasons provoking this situation. The first is the objectively existing shortage of specialists (Allyn, 2020), discrepancies between two expert interpreters (Sawan, 2017), and distrust of doctors, caused by the dissatisfaction of patients with existing medical care. For example, only 15% of Brazilians and 13% of Russians assess their medicine positively (Romir, 2020).

In this situation, the importance of the so called second opinion for patients increases dramatically (Benbassat, 2019), especially, a second opinion from a radiologist (re-analysis of diagnostic images, namely CT, MRI, radiography, mammography, PET-CT, etc.) or cardiologist (ECG, MRI of the heart, etc.).

In principle, such analysis can be carried out by an expert doctor remotely, using modern telecommunications, in combination with automated interpretation of high-tech medical information by means of artificial intelligence (Ahuja, 2019).
Besides, the high-tech medical information contained in patient requests could serve as a good basis for the formation of appropriate datasets, but for this it is necessary to develop appropriate organizational and motivational procedures.

Thus, there is the problem of integration of social media platforms and specialized web resources for the effective use of high-tech medical information, which is especially relevant for medicine in developing countries. In the article, this task is considered in relation to the Russian-speaking segment of the Internet.

Namely, we have experimentally studied the structure of public medical Internet communities typical for Russia. We found that they are characterized by self-organization. We have developed and launched a web resource for the effective use of high-tech medical information, and to form the motivational component of the resource, we use the identified structure of already existing network communities of medical focus. We use specialized chat bots as an effective means of integrating the developed resource and network communities.

2 BACKGROUND AND RELATED WORKS

Internet services, on which experienced doctors provide the patient with their reports on his high-tech medical information, are widely represented in medically developed countries, for example in USA – Second Opinions (from 295$), 2nd.MD ($3,000); in India – SeekMed (video communication with a specialist is provided). These services are websites or applications with a strictly defined business model, which implies the legitimacy and security of the transmitted information. Most of these services are expensive and therefore not available to everyone.

Similar services also exist in Russia, for example: Cardio-online; National Teleradiological Network. But, all of them are not free and have a very narrow specialization.

In addition to specialized Internet-services, there is also the possibility of interpreting high-tech medical information through social networks, like Facebook. In Russia and in most developing countries such as Iran, Malaysia etc., the Telegram messenger is much more popular. A large share of the Russian-speaking segment of the Internet is also occupied by the social network VKontakte. In these social networks, communities related to medicine, cardiology, cardiac surgery, radiology, and neurology are popular. But, although these communities are usually administered by professional doctors, it is impossible to guarantee the adequacy of the interpretation of medical information here.

The issues of obtaining and interpreting medical information in social media are studied in the literature mainly in the aspect of crowdsourcing (Wang, 2020; Kalantarian, 2019; Tucker, 2019). (McCoy, 2014) defines crowdsourcing to outsource a task to a group or community of people. (Tucker, 2019) concerns crowdsourcing activity as online collaboration systems.

Many studies suggest crowdsourcing to perform only separate, well-structured tasks - for example, for pre-clinical research (Tucker, 2019), for formatting incoming information, for improving the quality of the extracted facts (Kalantarian, 2019). To process information at a higher level by means of crowdsourcing, it is proposed to involve specialists. For example, in (Yoshida, 2016) hundreds of scientists were recruited first to generate, and then to assess competing health research ideas using a pre-defined set of priority-setting criteria. At the same time, there are examples of using crowdsourcing in artificial intelligence projects, most often related to annotation of medical data (Wang, 2020).

Insufficient attention is paid to the composition and structure of the interaction of crowdsourcing participants. As noted in the review (Créquit, 2018), crowd workers’ characteristics and crowdsourcing logistics are poorly reported in the reviewed articles. Crowd workers’ characteristics are frequently missing: even age and gender are not reported for about 60% of the studies.

Among the motivating factors for contribution or collaboration in medical crowdsourcing, various researchers distinguish recognition, curiosity, intrinsic satisfaction, or, in some situations, financial incentives (McCartney, 2013; Go, 2015; Chiauzzi, 2015). The (WHO, 2018) recommendations for underdeveloped countries suggest such an unexpected mechanism for motivating crowdsourcing, as the organization of challenge contests for health. However, in general this aspect of crowdsourcing remains outside the attention of researchers: according to the review (Créquit, 2018), of 202 studies motivations of crowd workers were recorded for 5 only.

The analysis performed allows us to draw the following conclusions.

The use of communities in social networks is convenient and accessible to all segments of the population, but at the same time it does not guarantee the legitimacy and reliability of the interpretation.
obtained. In such communities, there is "spontaneous crowdsourcing" supported by self-organization, but its use as a motivation mechanism has not been studied. On the other hand, existing specialized medical services have a high level of legitimacy and reliability, but most of them are expensive for users and do not provide for the use of incoming information as a source for the formation of datasets for use in artificial intelligence.

Therefore, the authors of this article set themselves the following tasks:

1. Explore the structure of organization and self-organization of participants in communities on social media platforms aimed at effective use of high-tech medical information.

2. Conduct a problem-oriented analysis of the needs and motivations of the participants in these communities.

3. Based on the analysis, develop a solution that provides integration of social media platforms and specialized web resources for the effective use of high-tech medical information.

3 METHODS AND MATERIALS


The detailed information about the community is given in the Appendix.

The main audience of the communities was determined after additional analysis of each community and detailed review of the participants. A detailed analysis of the content published in the communities showed that all communities have an administrator, who is a doctor, and community members can be roughly divided into three groups - patients, students, and professional doctors.

In our community research, we used only publicly available information. We used the Popsters analytics service and R programming language. We used the free developed environment RStudio, as well as the vkR package, which provides access to the VK API. To visualize the "friendship network" and build an interactive graph, we used the tkplot library.

For each VK community, we calculated 22 variables characterizing community users (Variables №1-9: Active Users, Population, Clear population, Members, Share Active Members, Connected Users, Connected Users Share, Isolates, Isolates Share), the “friendship networks” (Variables №10-19: Edges, Connected Components, Vertex Giant Component, Density, Modularity (from 0 to 1), Clusters, Mean Geodesics, Diameter, Mean Degree) and the structural indicators of community users (Variables №19-22: Female Share, Writer Share, Liker Share, Passive Share). The detailed description of the variables is presented in Appendix.

We used the k-means (MacQueen, 1967) and the LDA (Blei, 2003) algorithms to identify the community keywords.

To analyze messages in Telegram channels, we wrote a data grabber in the Python language using the telethon library. We used the analytics service and the grabber to see the total number of posts of a certain text length and certain content.

To view the dynamics of the visibility of the site Cardio-online, we used the resource Be1.ru.

4 RESULTS

Statistics on the number of subscribers to Communities №1 and 2 hosted on the VK platform are shown in Figure 1. It shows that the number of subscribers is growing in both communities for the last month.

For each VK community, we calculated 22 variables characterizing community users (Variables №1-9: Active Users, Population, Clear population, Members, Share Active Members, Connected Users, Connected Users Share, Isolates, Isolates Share), the “friendship networks” (Variables №10-19: Edges, Connected Components, Vertex Giant Component, Density, Modularity (from 0 to 1), Clusters, Mean Geodesics, Diameter, Mean Degree) and the structural indicators of community users (Variables №19-22: Female Share, Writer Share, Liker Share, Passive Share). The detailed description of the variables is presented in Appendix.

We used the k-means (MacQueen, 1967) and the LDA (Blei, 2003) algorithms to identify the community keywords.

To analyze messages in Telegram channels, we wrote a data grabber in the Python language using the telethon library. We used the analytics service and the grabber to see the total number of posts of a certain text length and certain content.

To view the dynamics of the visibility of the site Cardio-online, we used the resource Be1.ru.
Figure 2 shows Community №1 and №2 statistics for the total number of views for all posts. We can see a slight increase in each community in the total number of views for all publications. The growth has been most noticeable since the beginning of 2020, especially for Community №2.

Similar data for the communities №3 and №4 hosted on the Telegram platform are shown in Figure 3 showing active growth in subscribers. Figures 4, 5 display the total number of views in Telegram channels for all publications on different dates in the analyzed period. The Figures illustrate a significant increase of views in the period from 2019 to 2020 for all Telegram channels. The most pronounced growth has been observed since the beginning of 2020.

For comparison, Figure 6 illustrates dynamics of website «Almazov National Medical Research Centre of the Ministry of Health of the Russian Federation» visibility. The orange line shows the number of requests and the blue line shows the number of effective impressions. We can see that the number of requests (the frequency of entering keywords from site visibility in Yandex search per month, according to Wordstat) for the last month is 179. The number of site snippets shown to users in Yandex search results, according to the site visibility data is 1,816. The number of requests to the site decreases, the activity on the site is low. A summary of the obtained user indicators in Communities №1 and №2 under consideration is presented in Table 1.

<table>
<thead>
<tr>
<th>№</th>
<th>Variable</th>
<th>Com. №1</th>
<th>Com. №2</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Active Users</td>
<td>1,213</td>
<td>1,199</td>
<td>Users</td>
</tr>
<tr>
<td>2</td>
<td>Population</td>
<td>19,030</td>
<td>36,869</td>
<td>Users</td>
</tr>
<tr>
<td>3</td>
<td>Clear Population</td>
<td>18,392</td>
<td>35,936</td>
<td>Users</td>
</tr>
<tr>
<td>4</td>
<td>Members</td>
<td>18,228</td>
<td>35,936</td>
<td>Users</td>
</tr>
<tr>
<td>5</td>
<td>Share Active Members</td>
<td>7%</td>
<td>3%</td>
<td>% of group members</td>
</tr>
<tr>
<td>6</td>
<td>Connected Users</td>
<td>11,225</td>
<td>14,260</td>
<td>Users</td>
</tr>
<tr>
<td>7</td>
<td>Connected Users Share</td>
<td>63%</td>
<td>40%</td>
<td>% of group members</td>
</tr>
<tr>
<td>8</td>
<td>Isolates</td>
<td>7,167</td>
<td>21,676</td>
<td>Users</td>
</tr>
<tr>
<td>9</td>
<td>Isolates Share</td>
<td>39%</td>
<td>60%</td>
<td>% of group members</td>
</tr>
</tbody>
</table>

Based on the processed data, it can be concluded that users showing any activity on the pages of Community №1 and Community №2 constitute a small part of all users of the group, only 7% and 3%. At the same time, users who are on each other's friends list make up more than half of all members of Community №1 (63%) and less than half of all members of Community №2 (40%). There are also quite a few isolates in the communities.

A summary of the obtained structural indicators of the “friendship network” is presented in Table 2.
Table 2: Structural indicators of the “friendship network”.

<table>
<thead>
<tr>
<th>№</th>
<th>Variable</th>
<th>Com. №1</th>
<th>Com. №2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Edges</td>
<td>33,993</td>
<td>33,682</td>
</tr>
<tr>
<td>11</td>
<td>Connected Components</td>
<td>228</td>
<td>187</td>
</tr>
<tr>
<td>12</td>
<td>Vertex Giant Component</td>
<td>95%</td>
<td>96%</td>
</tr>
<tr>
<td>13</td>
<td>Density</td>
<td>0.0002</td>
<td>0.0005</td>
</tr>
<tr>
<td>14</td>
<td>Modularity</td>
<td>0.0216</td>
<td>0.0295</td>
</tr>
<tr>
<td>15</td>
<td>Clusters</td>
<td>264</td>
<td>234</td>
</tr>
<tr>
<td>16</td>
<td>Mean Geodesics</td>
<td>6.6</td>
<td>7.1</td>
</tr>
<tr>
<td>17</td>
<td>Diameter</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>18</td>
<td>Mean Degree</td>
<td>6.18</td>
<td>4.72</td>
</tr>
</tbody>
</table>

The general structure of the of the network of active users of Communities №1 and №2 is shown in Figure 7. The vertices represent users, next to the black circle, the user ID is indicated with blue numbers. An arc link is established between those group members who have each other in the friend list.

From the constructed graphs, we can see in more detail the structure of communities. From Figures 8-10, it is obvious that not all participants in communities have close ties - many participants are not connected with each other practically at all. There are many subgraphs in networks – clusters (Figure 6).

A summary of the community users’ structural indicators is presented in Table 3 (Community names have been translated from Russian). For a better understanding of the specificity of these communities’ profile and users’ interests, keywords were highlighted (Table 4). Keywords have been translated from the native language.

Table 3: Composition of community users.

<table>
<thead>
<tr>
<th>№</th>
<th>Variable</th>
<th>Com. №1</th>
<th>Com. №2</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Female Share</td>
<td>73.1%</td>
<td>70.2%</td>
</tr>
<tr>
<td>20</td>
<td>Writer Share</td>
<td>0.4%</td>
<td>0.1%</td>
</tr>
<tr>
<td>21</td>
<td>Liker Share</td>
<td>6.2%</td>
<td>3.2%</td>
</tr>
<tr>
<td>22</td>
<td>Passive Share</td>
<td>95.4%</td>
<td>96.9%</td>
</tr>
</tbody>
</table>

Table 4: Keywords of Communities №1 and №2.

<table>
<thead>
<tr>
<th>Keywords of Com. №1</th>
<th>Keywords of Com. №2</th>
</tr>
</thead>
<tbody>
<tr>
<td>monitoring</td>
<td>tests</td>
</tr>
<tr>
<td>Holter monitoring</td>
<td>medic ECG</td>
</tr>
<tr>
<td>analysis</td>
<td>subscribe</td>
</tr>
<tr>
<td>Holter examination</td>
<td>follow the news</td>
</tr>
<tr>
<td>ECG</td>
<td>participate in public life</td>
</tr>
<tr>
<td>patient risk</td>
<td>get to know ECG</td>
</tr>
<tr>
<td>disease</td>
<td>vascular</td>
</tr>
<tr>
<td>help</td>
<td>cardiologist</td>
</tr>
<tr>
<td>year</td>
<td>artery</td>
</tr>
<tr>
<td>heart</td>
<td>stenosis</td>
</tr>
<tr>
<td>female</td>
<td>coronary</td>
</tr>
<tr>
<td></td>
<td>literature</td>
</tr>
<tr>
<td></td>
<td>aorta</td>
</tr>
</tbody>
</table>

From Table 4, we can see that Community №1 aims to provide high-tech health information assessment and advice on Holter monitoring, and
Community №2 aims to educate and deepen knowledge in cardiology.

Figures 11, 12 show the total number of posts for Communities №3-6. Left to Right Values: small (less than 160 symbols), average (from 160 to 1000) and large (more than 1000) length.

Figures 13, 14 display the total number of publications with a certain content. Left to Right Values: text, photos, videos, links.

5 DISCUSSION

Since the subject of the medical Telegram channels and VK communities under consideration is rather narrow and specific, the number of subscribers is not large. However, from the Figures 1, 3 we can see that the number of subscribers is growing significantly, especially from the second half of 2020. This is mostly pronounced for the Telegram channels. Figures 2, 4, 5 illustrate the increase of the total number of views for all publications. This indicates an increase in the level of people's interest in these medical topics.

The obtained indicators of users of Community № 1 (Table 1) indicate a small percentage of active users (7%) and many isolates (7,1 67). From the structural indicators of the “friendship network” (Table 2) and the general network structure, we can see that not all users are linked with each other. It shows the diversity of the audience, which suggests an individual approach to every person. From the obtained structural indicators of users of Community № 1 (Table 3), it can be seen that a very small part of the group creates content and a huge part of the group does not show any activity. This allows us to say that most of the community members are not aimed at active communication among most participants.

Having additionally analyzed the participants and their actions in the communities, we made sure that users enter these communities mainly to receive an assessment of high-tech medical information (Table 4), which is provided by a limited circle of people (doctors). At the same time, medical students and doctors join these communities to gain experience and deepen their knowledge (Figure 10).

Table 4 shows the thematic profile of communities, their interests and needs. In combination with the analysis of the group structure (Figures 7, 9), this once again confirms that people have little horizontal connection with each other.

Figure 10 shows the most common network clusters are formed by a group of students from one medical university and a group of medical workers. However, there are several leaders with great activity – they are often contacted, this can be seen from the variables and pictures.

Figures 11, 12 show that most publications in Communities № 3 and №4 have medium and large length. This is due to the patients' description of their symptoms and the first medical opinion, as well as detailed answers from doctors and medical students. A more detailed analysis of the posts shows that people write their medical data directly. This means that the need for additional interpretation of their
medical data is so great that it often outweighs the requirements for privacy and legitimating. This circumstance should be taken into account when building a web resource.

Figures 13, 14 show that users send their data in various graphical formats to get an assessment of high-tech medical information. At the same time, the radiology Community №5 is characterized by short messages. This is because videos and links to useful materials are more often published on the main channel. In Community №6 there are more long messages.

Summing up the results of the analysis, we can state that in order to obtain an effective assessment of high-tech medical information, residents of Russia address primarily in public Internet spaces. There emerge and become popular communities working in the mode of “spontaneous crowdsourcing”, and they have a pronounced self-organization. Users are divided into groups - patients, students and doctors - with their interests that do not contradict each other and are fully satisfied there. This allows such communities to provide the necessary informational services to their participants for free.

However, the medical information discussed in such communities is transmitted and disseminated in an open manner, without regard to privacy and legitimating requirements. In addition, despite the obvious value of this information, it is practically not used as a source for the formation of datasets for subsequent use in electronic medicine through artificial intelligence.

It points to the relevance of creating a publicly available solution for evaluating high-tech information, combining the accessibility and ease of communication inherent in social networks, providing a legitimate second opinion for patients and at the same time allowing the use of the information provided to form datasets. The authors see such a solution as a combination of a specialized network resource and a set of chatbots that integrate it with network platforms. The results presented above made it possible to formulate the requirements for the proposed solution:

1. VK communities and Telegram channels are not legitimate enough. The chatbot solves this problem because it integrates with a high-tech service.
2. People need to be confident that a service can be trusted. Therefore, when designing a chatbot, it is necessary to ensure the concealment of personal data.
3. Users need the ability to provide data in various graphic formats for convenience.

4. The opportunity to exchange information in a simple dialogue form is necessary. It is necessary to design different dialog scenarios, different levels of description for different clusters of users, considering their peculiarities. Therefore, it is necessary to foresee several chatbots targeted at specific user groups.

6 PROPOSED SOLUTION

A use-case diagram of the developed specialized web resource is shown in Figure 15. The resource works on the principle of teleradiology. Its datasets are formed thanks to the attending physicians and their patients, who are ready to provide their anonymized medical information for forming datasets. The main actors of this resource are the patient, the expert doctor, the trainee, and the owner of the resource: each of them gets a certain benefit from the use. It also has a high level of privacy and security: this inspires trust among users. The resource allows using high-tech tools: it served as the motivation for creating a bot. The chatbot is served as an integration tool. Through its implementation in social networks, it engages people in using of a high-tech tool and brings them together, thereby meeting their needs. The chatbot is easy to use and does not require any special knowledge.

Image: Use Case diagram of the specialized web resource.

Figure 16 shows a use case diagram describing the integration of the specialized web resources with social media platforms. The interaction of social platforms users with the service is realized through the mechanism of chatbots. Our chatbots (TelegramBot-MCP and VKBot-MCP) are developed individually for each platform, since the mechanism of interaction between a chatbot and a social platform depends on the API offered by the platform.

The business logic of interaction between the user and the bot is implemented on the side of the bot's
software module. Social platforms provide users with a conversational interface to a chatbot in a familiar environment. Through the chatbot, the user can send a request for a second opinion to the Platform and receive answer as soon as it is ready.

The interaction of chatbots and the main part of the platform is carried out through a special API, which allows to scale the system and transparently connect other social platforms. Currently the specialized web resource for the effective use of high-tech medical information «ITMO University's Medical Crowdsourcing Platform» (ITMO_MCP) has been launched on the platform of the ITMO University (St. Petersburg, Russia).

![Figure 16: Use Case diagram.](image)

### 7 CONCLUSION

The article discusses the possibility of creating a public resource for the effective use of high-tech medical information. The tasks set by the authors of the article have been successfully fulfilled. We have investigated the structure of organization and self-organization of community members in social networks aimed at obtaining and effectively using high-tech medical information. We have also analyzed public medical Internet spaces (VK medical communities in cardiology and Telegram medical channels in cardiology and radiology), with the help of which the motivations and needs of their participants were identified.

Based on the results of the analysis of medical communities, we proposed a solution that provides the integration of social media platforms and specialized web resources for the effective use of high-tech medical information.

### ACKNOWLEDGEMENTS

This work was supported Grant of the President of the Russian Federation for state support of young Russian scientists - candidates of science, MK-5723.2021.1.6.

### REFERENCES


Chiauzzi E., Eichler G., Wicks P. Crowdsourcing Advancements in Health Care Research. December 2016. DOI: 10.1016/B978-0-12-802115-6.00016-1


