

MODEL OF IDENTIFICATION OF OPINION LEADERS IN TASKS OF SOCIAL NETWORKS MONITORING

A. Chudakova, Zhongdong Wu, M. Smirnova

Lanzhou Jiaotong University

730070, China, Gansu Province, Lanzhou City, Anning District, Anning West Road, 88. E-mail: alliks1@gmail.com

Purpose. To improve the existing approaches of searching for opinion leaders in the tasks of social networks monitoring by creating a model and the software tool to automatize this process and visualize the result for the end user (based on QZone social network). **Methodology.** When solving the set tasks, the authors have used general scientific methods such as analytical research of existing methods and software tools for the identification of opinion leaders in social networks, existing approaches of social networks data mining, and existing approaches of Chinese text preprocessing; synthesis of a model for opinion leaders' identification. **Findings.** The question of opinion leaders' identification by programming methods has been considered. Opinion leader categories have been defined. The criteria for identification of opinion leaders have been defined and properly analyzed. The set of conditions for identification of opinion leaders by programming methods have been synthesized. This will allow us to develop a software tool for the given task. As a result, the authors have improved existing approaches of identification of opinion leaders in the tasks of social networks monitoring (based on Chinese social network QZone) by developing a model that can be used by software application that represents search results for the end user in conditions close to real time and provides a set of tools to apply a set of requirements. **Originality.** Nowadays, an identification of opinion leaders is an actual task and there is no one universal method that can solve it. The authors have proposed a method to solving this problem by processing basic information about social networks users. **Practical value.** The research represents a method that can be used by western clients to analyse Chinese society for marketing purposes, and, with minor modifications, can be used to analyse any other social network. References 10, figures 7.

Key words: opinion leaders, social networks analysis, intellectual text analysis, big data.

МОДЕЛЬ ПОШУКУ ЛІДЕРІВ ДУМОК В ЗАДАЧАХ МОНІТОРИНГУ СОЦІАЛЬНИХ МЕРЕЖ

A. Чудакова, Чжундун У, М. Смірнова

Ланьчжоуський Транспортний Університет

730070, Китай, провінція Ганьсу, м. Ланьчжоу, район Аньнін, вул. Аньнін Вест, 88. E-mail: alliks1@gmail.com

Удосконалено існуючі методи пошуку лідерів думок у задачах моніторингу соціальних мереж шляхом розробки моделі та програмного засобу для автоматизації цього процесу та візуалізації результату для кінцевого користувача (на основі соціальної мережі QZone). У процесі вирішення поставленої задачі проведено порівняльний аналіз існуючих методів та програмних засобів для пошуку та виявлення лідерів думок в соціальних мережах, а також методів обробки тексту (китайська письменність), розроблено модель пошуку лідерів думок. Розглянуто питання пошуку лідерів думок в задачах моніторингу соціальних мереж. Визначено категорії лідерів думок. Визначено критерії пошуку лідерів думок та проведено їх детальний аналіз. Визначено набір вимог для ідентифікації лідерів думок програмними методами, що дозволяє розробити програмне забезпечення для вирішення поставленої задачі. Як результат, було вдосконалено існуючі методи пошуку лідерів думок в задачах моніторингу соціальних мереж (на основі соціальної мережі QZone) шляхом створення моделі, що може використовуватись при розробці програмного забезпечення для представлення результатів пошуку кінцевому користувачу в умовах наближених до реального часу. В наш час визначення лідерів думок є актуальною задачею та не існує універсального методу її вирішення, таким чином оригінальність роботи полягає в розробці методу вирішення даної проблеми шляхом обробки базової інформації користувачів соціальних мереж. Представлена модель може використовуватись західними клієнтами для аналізу китайського суспільства в соціальній мережі QZone з точки зору маркетингу, а також, з незначними модифікаціями, для пошуку лідерів думок в інших соціальних мережах.

Ключові слова: лідери думок, аналіз соціальних мереж, інтелектуальний аналіз тексту, великі дані.

PROBLEM STATEMENT. The concept of a social network was used by sociologists in the 1920s to study the interconnections between participants of different communities. Modern social networks have significantly changed the formulation of the issue - today researchers have a «free» resource for surveys, and the rapid spread of social online services and the development of Big Data technologies have raised an interest in using information from social networks in various industries. Combined usage of structural and content data potentially allows using social networks to solve a wide range of business tasks: fraud fighting, brand management, advertising of goods and services, the formation of new distribution channels etc.

Social networks monitoring is a selection of users' texts in social networks, based on certain criteria as well as the process of collection this selection and its interpretation.

There are several basic marketing tasks that are effectively solved by monitoring, such as: early detection of negative comments; competitors' analysis; evaluation of promotion effectiveness; getting feedback; search for thematic discussions; search for opinion leaders etc.

Opinion leader is the person who can make an influence on other people opinions. Usually they are well-known experts or bloggers as well as celebrities, famous athletes - in general, people who have achieved good results.

Nowadays China is a huge market for foreign goods with millions of potential customers. Companies are already well aware that through Weibo or WeChat they can effectively sell their products and services. Therefore, search of opinion leaders in Chinese social networks is an actual task, considering that there are not a lot of tools that can be used by non-Chinese speakers.

The research aims to improve the existing approaches of searching for opinion leaders and develop the software tool to automatize this process and visualize the result for the end user (based on QZone social network).

EXPERIMENTAL PART AND RESULTS OBTAINED. In China there are 649 million Internet users, 80 % of them regularly use social networks [1], spending online more than 25 hours a week. 38 % of Chinese consumers make purchasing decisions based on recommendations in social networks, while more than 75 % write a review about the purchased item in online communities at least once a month. For comparison - in the USA this number is less than 20 %. Obviously, the brands online presence in China is important.

Chinese Internet is essentially controlled by the government, international players such as Facebook, WhatsApp, Instagram and other services are prohibited and replaced by local analogues WeChat, Weibo, QQ, Baidu Tieba and so on.

Despite the fact that Chinese almost completely ignore the western infrastructure of social networks and search engines, protecting local developers, today they have managed to create their own digital ecosystem, which in scale and scope of the audience sometimes surpasses Western analogues. For example:

- Tencent QQ (messenger) – 830 million users (more

than the audience of LinkedIn, Facebook and Twitter combined);

- QZone (MySpace analogue) – 755 million users;
- Sina Weibo (platform for microblogging, Twitter analogue) – 600 million users;
- WeChat (messenger) – 468 million users;
- Pengyou (social network) – 260 million users;
- Renren (similar to Facebook) – 195 million users.

To imagine the fact that you can write a post on Weibo and it can potentially be read by millions, and even tens of millions of users is quite amazing.

The online market in China is still very fragmented. It is not easy for small and medium Western companies to enter the market; therefore, it is necessary to find competent opinion leaders to advertise goods and products in the Chinese market.

In the market with low brand loyalty and actual knowledge about them, the opinion leaders can quickly and effectively help to get the customers. Of course, this approach is widely used in the West, but in China it has a completely different scale – according to the significance of social networks in the life of Chinese people.

For example, Chinese actress Fan Bingbing alone has made a profit of \$75 billion for Taobao network. One of the most influential fashion bloggers Gogoboi has a base of 4.43 million subscribers and, accordingly, great potential for stimulating sales of brands and products that he mentions in his blog. Gogoboi is attracted by various brands, for example, Louis Vuitton.

Fig. 1 represents an example of using a celebrity Weibo page for advertising a product. Thanks to a huge fanbase a simple picture with an ice-cream got 210 thousand reposts, 50 thousand comments and 520 thousand likes.



Figure 1 – Usage of celebrity Weibo page (as opinion leader) for an advertisement

Unfortunately, opinion leaders of this level often cost a huge amount of money. However, in addition to

the stars of any category (travel, beauty, education of children and others), China has its own leaders of pub-

lic opinion. Their participation allows you to quickly and fairly cheap achieve the desired results.

It is important to choose a reliable person with the necessary brand reputation, which is able to produce high-quality content in various social media.

Chinese netizens actively engage in discussions that could directly affect companies. A recent study by Ogilvy One in China found that 55 percent of China's netizens had initiated or participated in online discussions about companies. Therefore, to find opinion leaders among them is an important and difficult task.

SINA's microblogging site Weibo has launched an open platform to help key opinion leaders connect with online vendors and service providers as it intends to help boost development in the online marketing ecosystem [6]. Weibo is already working with more than 2,000 online vendors and 15,000 online celebrities to help them interact with followers.

To consider the fact that other Chinese social networks don't have such useful function, we decided to search for opinion leaders in QZone. It is the second most popular social network in China and has a desktop version. It allows users to write blogs, keep diaries, send photos, listen to music, and watch videos.

In our research, we decided to pay main attention to the problem of finding opinion leaders rather than data collection problem, therefore, we have chosen the easiest way to collect data – to use the services of specialized companies that collect and constantly update data from a variety of sources. The main advantage here is the speed of obtaining information, which is essential for large volumes of the client base and the use of various social networks. The disadvantage is a paid subscription for data updates.

For convenience, it makes sense to divide opinion leaders by the number of connections they can establish with others, i.e. on the intensity of communication. Here is a simple classification:

- ordinary leaders (common people, but they act as sources of information about a certain category of goods and are able to influence the choice of people in their circle of communication);
- mega-leaders, or celebrities (persons who have achieved certain results and are known to a wide range of people);
- experts (have an excellent knowledge in a specific field);
- social leaders (earned the trust of others, thanks to charisma and social activity).

Detection of social leaders is a complicated process. After reviewing the existing approaches of searching for opinion leaders, we defined the criteria, that help to find such people.

- Number of followers. It is the most significant criterion, however, usually there are not only ordinary followers, but also bots, advertisement pages etc. We have selected 200 users with over a half of million followers for analysis. The results are shown in Fig. 2.

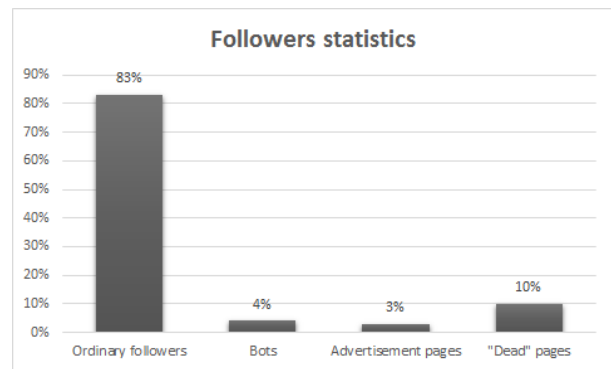


Figure 2 – Followers statistic

Therefore, the number of real followers is counted as shown in formula (1):

$$Fn = Fn_{pre} \cdot 83\%, \quad (1)$$

where Fn – number of real followers, Fn_{pre} – the whole number of followers, 83% – the number of ordinary followers among the all followers.

But a large number of subscribers is not a sufficiently accurate indicator to decide if the user is the opinion leader or not. The number of published messages is also important (consider that they are written by the user themselves, and not reposted). In addition, it is necessary to take into account the reaction of the followers in the form of the number of comments to the message, the number of likes and reposts.

- Posts (Pn). The research has shown that it is a wrong decision to take into account all posts starting from the account creation (except celebrities, other categories of opinion leaders don't get their followers in a blink of eye) and we aimed at searching for opinion leaders that has spent some time earning the posts that respect, therefore we take in consideration the posts that are one year old and newer. Therefore, Pn – is a number of user's own posts (not reposts) within a year.

- «Likes». We used the same selection of 200 users to calculate the involvement rate for «likes» in Chinese social networks (Lr). We defined that for opinion leaders each post should have the number of «likes» that is not less than 2,4% of the subscribers. For example: if the user has 20 thousand subscribers, then one post should have at least 480 «likes». The results of selection analysis are given at Fig. 3.

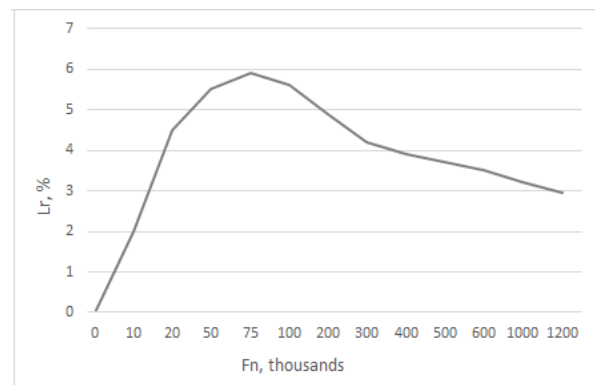


Figure 3 – Relation of «likes» involvement rate to the number of followers

But there is the trend: the bigger the number of subscribers, the lower the level of involvement and vice versa. Therefore, we use the formulas (2) and (3) to calculate the involvement rate for each user.

$$Lr = \frac{Ln_{av}}{Fn} \cdot 100\% , \quad (2)$$

$$Ln_{av} = \frac{\sum_{i=1}^{Pn} Ln_i}{Pn} , \quad (3)$$

where Lr – «likes» involvement rate; Ln_{av} – an average number of «likes» at posts (within one last year); Ln_i – number of likes at post i .

This way we are looking for accounts that meet both of the following requirements:

$$\begin{cases} Fn \geq 10000 \\ Lr \geq 2\% \end{cases} \quad (4)$$

• Comments. The calculations on the comments involvement rate is similar to the calculations on the «likes» involvement rate and are represented with formulas (5) and (6). The results of selection analysis are given at Fig. 4.

$$Cr = \frac{Cn_{av}}{Fn} \cdot 100\% , \quad (5)$$

$$Cn_{av} = \frac{\sum_{i=1}^{Pn} Cn_i}{Pn} , \quad (6)$$

where Cr – comments involvement rate; Cn_{av} – an average number of comments at posts (within one last year); Cn_i – number of comments at post i .

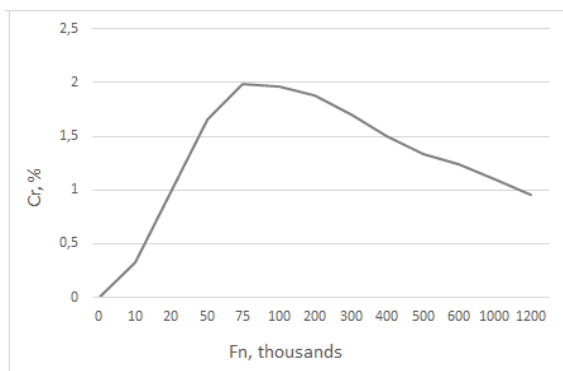


Figure 4 – Relation of comments involvement rate to the number of followers

We should consider that the involvement rate for comments will always be smaller than the involvement rate for «likes» and our research has shown that normally opinion leaders' posts have number of comments that is about 0,7% of followers' number (fig. 4). This way we are looking for accounts that meet the following requirements:

$$\begin{cases} Fn \geq 5000 \\ Cr \geq 0.7\% \end{cases} \quad (7)$$

• Reposts. The results obtained during research on the involvement ratio for reposts has shown that QZone is Twitter-like social network, where the number of reposts usually exceed the number of likes (unlike in Facebook and VKontakte). The results of selection analysis are given at Fig. 5. This way we come to formulas (8) and (9):

$$Rr = \frac{Rn_{av}}{Fn} \cdot 100\% \quad (8)$$

$$Rn_{av} = \frac{\sum_{i=1}^{Pn} Rn_i}{Pn} , \quad (9)$$

where Rr – reposts involvement rate; Rn_{av} – an average number of reposts; Rn_i – number of reposts of post № i .

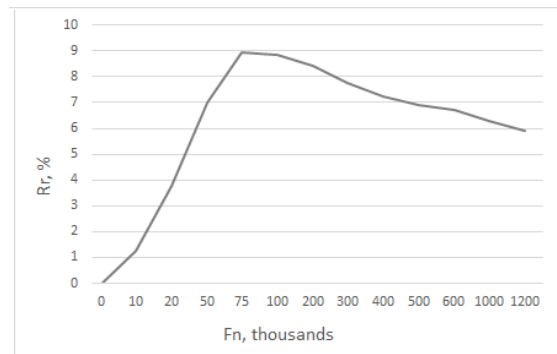


Figure 5 – Relation of reposts involvement rate to the number of followers

Therefore, we are looking for accounts that meet the following requirements:

$$\begin{cases} Fn \geq 5000 \\ Rr \geq 5\% \end{cases} \quad (10)$$

• Number of posts per week. The number of posts within a year is used to evaluate involvements rates, but number of posts per week better characterize opinion leaders. Usually such people have at least 5 posts per week. If there were no posts for a week, or even a month that this person doesn't meet the requirements our research aimed at. Therefore, $Pw \geq 5$, where Pw – is number of posts per week.

• Opinion leader category. Earlier we divided opinion leaders into 4 categories. It is difficult to divide ordinary leaders, experts and social leaders by programming methods, but basing on followers' number we can divide «celebrities» category. As their services often cost a huge amount of money, it might come in hand to exclude them from search results, if there is a need.

First of all, we want to notice that in China there is an enormous number of social networks users, therefore Chinese celebrities usually have a huge number of followers, that is counted in tens of millions.

Secondly, most of Chinese celebrities use Weibo to communicate with their fans and barely appear at QZone.

By analyzing the list of celebrities that have their pages at QZone and the number of their followers, we

came to decision, that users who have over 5 million followers can be tagged as «celebrities».

- Field of activity. It is not enough to define a list of opinion leaders, but we need to know the field of their activity as well, because we can't ask a person to advertise cosmetics while he/she is giving feedbacks for new mobile devices. And that's where we can apply semantic analysis.

We need to clean the posts up before doing any analysis. This process involves removing content, such as punctuation (has no meaning), and we need to remove any content that causes errors:

- convert to lower case (no need to do for Chinese content);
- remove punctuation;
- remove control characters (or non-printing characters);
- remove numbers;
- remove stop words (ex. the, is, at, which, and so on);
- remove html-links;

- remove tags (optional, should be done at the beginning of text cleaning).

Here is a short list of R functions that have been used to achieve desired results:

```
p=gsub('[:punct:]',",p)
p=gsub('[:cntrl:]',",p)
p=gsub('\\d+',",p)
p=gsub('[:digit:]',",p)
p=gsub('#\\w+',",p)
p=gsub('http\\w+',",p)
p=gsub("^\\s+|\\s+$", "", p)
```

Furthermore, packages «chinese.misc» and «translateR» were used to define words in Chinese and translate them into English, so we can define top 10 most used words at user's wall.

Fig. 6 represents a sample extraction of 5 users from DB that meet requirements, given above.

Considering all the above we can define the model of identification of opinion leaders in tasks of social networks monitoring, basing on Chinese social network QZone, that is represented in Fig. 7.

SN	QZid	Sex	Fn	Lr	Cr	Rr	Words
新浪娱乐	1530194583	f	5783	4,7	2	9	learn homework student social classroom solution easy lector key task
包爽	125092995	m	24985	4,3	1,4	6,2	green recycle energy eco Earth ecology solution nature global organic
搞笑百科	291753045	m	182578	3,5	1,4	7,4	computer laptop camera software user application menu different allow features
CCTV新闻	383771753	f	694176	3,7	2	7,8	cctv news public information world report research question impact rise
苏有朋	622009404	m	2548000	2,6	0,9	5,4	fan miss hug activity photo beautiful thank meet movie music

Figure 6 – A sample extraction from DB

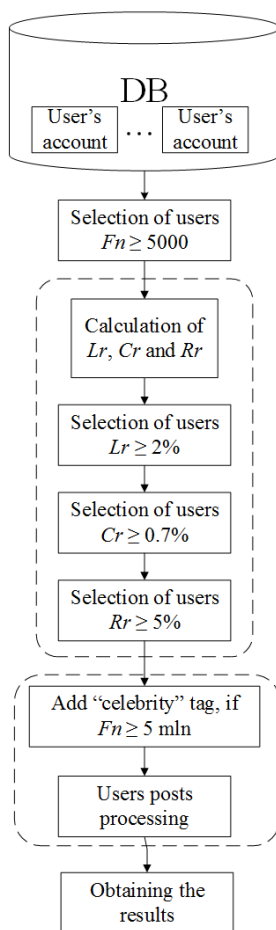


Figure 7 – The model of identification of opinion leaders

The results obtained during experiments have proved the efficiency of developed model.

- Plans for further research. Firstly, we are working on ways of improvement of developed model. Secondly, there is a planning process on the development of software, which will use the proposed model to find opinion leaders in the social network QZone. Functionality and the interface (convenient for the end user) are under discussion.

CONCLUSIONS. The research analyzes various approaches of opinion leaders' identification and, as result, suggests the model of identification of opinion leaders in tasks of social networks monitoring. There are few existing methods that can solve given task and our research is a one step forward in this field.

The experiments show that the suggested model decreases time of obtaining results significantly due to storing the initial data and search results on the computing device in DB.

In the future research, we consider the development of software application that can be used by western clients to analyse Chinese society for marketing purposes.

The suggested model can be applied to other social networks but with the correction of the peculiarities of these social networks, their restrictions and their data parameters.

ACKNOWLEDGEMENTS. We would like to thank Rustam Inoiatov, Oksana Chudakova and Anna Judina, for their moral support and wonderful contribution towards completion of this research.

REFERENCES

1. Mildberry. *5 Ways to Conquer the Chinese User*. Available at: http://www.mildberry.ru/cases/china_consumer/ (accessed April 02, 2017).
2. Murzin F. A. (2015), "A software package for a social network data analysis", *International Journal "Programmnye produkty i sistemy"* № 4, Tver, Russian Federation, 2015, pp. 188–197.
3. Jie Yang, Yecies B. (2016), "Mining Chinese social media UGC: a big-data framework for analyzing Douban movie reviews", *"Journal of Big Data"*, Springer, January, 2016.
4. King C. (2004), *A Survey on Text Mining in Social Networks*, Cambridge University Press, Cambridge, United Kingdom.
5. The R Project for Statistical Computing. Available at: <https://www.r-project.org/> (accessed April 02, 2017).
6. Ding Yining *Weibo sets up open platform to connect opinion leaders with online vendors*. Available at: <http://www.shanghaidaily.com/business/Weibo-sets-up-open-platform-to-connect-opinion-leaders-with-online-vendors/shdaily.shtml> (accessed April 02, 2017).
7. Goddard C. (2011), *Semantic Analysis: A Practical Introduction*, Oxford University Press, Oxford, United Kingdom.
8. Landauer T., McNamara D., Dennis S., Kintsch W. (2014), *Handbook of Latent Semantic Analysis*, Psychology Press, Hove, United Kingdom.
9. G. van Heerden (2009), *Finding and utilizing opinion leaders: Social networks and the power of relationships*, *S.Afr.J.Bus.Manage*, 2009, 40(3), Sweden.
10. Katona Z. (2012), *How to Identify Influence Leaders in Social Media*. Available at: <https://www.bloomberg.com/view/articles/2012-02-27/how-to-identify-influence-leaders-in-social-media-zsolt-katona> (accessed April 02, 2017).
11. Russell M. (2013), *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*, O'Reilly Media, Sebastopol, California, USA.

МОДЕЛЬ ПОИСКА ЛИДЕРОВ МНЕНИЙ В ЗАДАЧАХ МОНИТОРИНГА СОЦИАЛЬНЫХ СЕТЕЙ**А. Чудакова, Чжундун У, М. Смирнова**

Ланьчжоуский Транспортный Университет

730070, Китай, провинция Ганьсу, г. Ланьчжоу, район Аньнин, ул. Аньнин Вест, 88. E-mail: alliks1@gmail.com

Усовершенствованы существующие методы поиска лидеров мнений в задачах мониторинга социальных сетей путем разработки модели, а также программного обеспечения для автоматизации этого процесса и визуализации результата для конечного пользователя (на основе социальной сети QZone). В процессе решения поставленной задачи проведен сравнительный анализ существующих методов и программного обеспечения для поиска и выявления лидеров мнений в социальных сетях, а также методов обработки текста (китайская письменность), разработана модель поиска лидеров мнений. Рассмотрен вопрос поиска лидера мнений в задачах мониторинга социальных сетей. Определены категории лидеров мнений. Определены критерии поиска лидеров мнений и проведен их детальный анализ. Определен набор условий для идентификации лидеров мнений программными методами, что позволяет разработать программное обеспечение для решения поставленной задачи. Как результат, были улучшены существующие методы поиска лидеров мнений в задачах мониторинга социальных сетей (на основе социальной сети QZone) путем разработки модели, которая может использоваться при создании программного обеспечения для представления результатов поиска конечному пользователю. В наше время определение лидеров мнений является актуальной задачей и не существует универсального метода ее решения, таким образом актуальной работы состоит в разработке метода решения данной проблемы путем обработки базовой информации пользователей социальных сетей. Представленная модель может использоваться западными клиентами для анализа китайского общества в социальной сети QZone с точки зрения маркетинга, а также, с небольшими изменениями, для поиска лидеров мнений в других социальных сетях

Ключевые слова: лидеры мнений, анализ социальных сетей, интеллектуальный анализ текста, большие данные.

Стаття надійшла 10.10.2016.